GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:49:18 ; Search time 48 Seconds
(without alignments)
2195.718 Million cell updates/sec

Title: US-09-590-991-6
Perfect score: 4398
Sequence: 1 YPVVLADTSSSEDALNISDK.....PGKNIKSTLNVINGKSTYGY 664

Scoring table: BLOSUM30
Gapop 8.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4398	100.0	664	22	AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21	AAY81710	Streptococcus pneu
3	4380.5	99.6	2140	24	ABU01020	S. pneumoniae type
4	4171	94.8	1007	19	AAW61246	Streptococcus pneu
5	4171	94.8	1007	23	ABP54664	S. pneumoniae SP12
6	975.5	22.2	192	19	AAY86159	S. pneumoniae deri
7	962.5	21.9	1962	12	AAR10559	Mutant protease (A
8	954.5	21.7	1968	12	AAR10941	Mutant protease (d
9	952.5	21.7	1962	12	AAR10557	Mutant protease (A
10	952.5	21.7	1974	12	AAR10940	Mutant protease (d
11	951.5	21.6	1962	12	AAR10558	Mutant protease (A
12	948.5	21.6	1962	12	AAR10560	Mutant protease (K
13	948.5	21.6	1962	12	AAR10561	Mutant protease (N
14	948.5	21.6	1962	12	AAR10563	Mutant protease (K
15	947	21.5	1959	12	AAR10562	Mutant protease (d
16	938.5	21.3	1560	10	AAP94145	S. cremoris protei
17	896.5	20.4	1946	19	AAW47273	Lactobacillus bulg
18	875	19.9	1570	23	ABP29894	Streptococcus poly
19	875	19.9	1570	23	ABP30531	Streptococcus poly
20	875	19.9	1590	23	ABP28560	Streptococcus poly
21	873	19.8	1550	23	ABP30895	Streptococcus poly
22	869.5	19.8	885	23	ABP27332	Streptococcus poly
23	868	19.7	885	23	ABP29796	Streptococcus poly
24	867.5	19.7	1239	23	ABP25822	Streptococcus poly
25	865.5	19.7	1579	24	ABP71293	Group B Streptococ
26	864	19.6	1233	23	ABP30203	Streptococcus poly
27	862	19.6	1233	23	ABP29675	Streptococcus poly
28	857	19.5	1647	23	ABP28561	Streptococcus poly
29	797.5	18.1	690	11	AAR04904	Residual protease-
30	792	18.0	2710	17	AAR95016	C. difficile toxin
31	792	18.0	2710	19	AAW68387	Clostridium diffic
32	787.5	17.9	690	18	AAW16337	Bacillus subtilis
33	787.5	17.9	690	18	AAW01787	Residual protease
34	782	17.8	806	13	AAR27481	RP-III residual pr
35	780.5	17.7	2835	23	ABB98574	Dextran saccharase
36	774	17.6	815	23	ABB91806	Herbicidally activ
37	771	17.5	922	19	AAW37372	Trimmed enzyme pro
38	769	17.5	1475	11	AAR08221	Recombinant alpha
39	764.5	17.4	1582	23	ABB47307	Listeria monocytog
40	761.5	17.3	903	17	AAR87007	Hyperthermostable
41	761.5	17.3	1398	17	AAR87008	Protease. Pyrococ
42	761.5	17.3	1398	18	AAW24124	Pyrococcus furiosu
43	761.5	17.3	1398	20	AAW94839	WO9856926 Seq ID 6
44	761	17.3	1941	23	ABP65509	Bifidobacterium lo
45	760.5	17.3	1185	13	AAR22675	Collagen binding p

ALIGNMENTS

RESULT 1
AAB48342
ID AAB48342 standard; Protein; 664 AA.
XX
AC AAB48342;
XX
DT 20-APR-2001 (first entry)
XX
DE S. pneumoniae Sp128 polypeptide.
XX
KW Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KW antibacterial; auditory; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO200076540-A2.
XX
PD 21-DEC-2000.
XX
PF 09-JUN-2000; 2000WO-US15925.
XX
PR 10-JUN-1999; 99US-0138453.
XX
PA (MEDI-) MED IMMUNE INC.
XX
PI Adamou JE, Choi GH;
XX
DR WPI; 2001-112197/12.
DR N-PSDB; AAC84741.
XX
PT New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT blood infections -
XX
PS Claim 4; Page 47-50; 54pp; English.
XX
CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Sp128 polypeptide.
XX
SQ Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 4e-64;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180

Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

RESULT 2
AAY81710
ID AAY81710 standard; Protein; 2120 AA.
XX
AC AAY81710;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID3.
XX
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006738-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB02452.
XX
PR 27-JUL-1998; 98GB-0016336.
PR 19-MAR-1999; 99US-0125329.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
DR WPI; 2000-195301/17.
DR N-PSDB; AAZ91806.
XX
PT Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections -
XX
PS Claim 2; Page 41-42; 76pp; English.
XX
CC This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments)are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis.
XX
SQ Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 9.6e-63;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
```

```
Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

```
RESULT 3
ABU01020
ID   ABU01020 standard; Protein; 2140 AA.
XX
AC   ABU01020;
XX
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #590.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media;
KW   ear infection; antiinflammatory; antibacterial; immunostimulant;
KW   auditory; respiratory; gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB02163.
XX
PR   27-MAR-2001; 2001GB-0007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX06302.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT   or ear infection -
XX
PS   Claim 1; SEQ ID No 1180; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the
CC   proteins, treating a patient by administering the protein, DNA or
CC   antibody (in a composition), a kit comprising first and second primers,
CC   which are the nucleic acid cited above or fragments between nucleotides
CC   8-100 of a sequence not defined in the specification, for amplifying a
CC   target sequence contained within a Streptococcus nucleic acid sequence,
CC   where the first primer is substantially complementary to the target
CC   sequence and the second primer is substantially complementary to the
CC   complement of the target sequence, and where the parts of the primers
CC   having substantial complementarity define the termini of the target
CC   sequence to be amplified, assay comprising contacting a test compound
CC   with the protein, and determining whether the test compound binds to the
CC   protein and a Streptococcus  pneumoniae bacterium, where one or more
CC   genes encoding the proteins has been rendered inactive. The proteins,
CC   nucleic acid molecules, antibody and compositions are useful as
CC   medicaments for treating or preventing a disease or infection due to
CC   streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC   sepsis, otitis media or ear infection. They are also useful in developing
CC   vaccines, diagnostics and antibiotics. The methods are useful for
CC   identifying immunodominant proteins. The present sequence is one of
CC   the 2469 proteins expressed by the identified coding regions from the
CC   genomic sequence.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2140 AA;
```

```
  Query Match             99.6%;  Score 4380.5;  DB 24;  Length 2140;
  Best Local Similarity   99.6%;  Pred. No. 1.9e-62;
  Matches  664;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    2;

Qy          1 Y-P--VVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 57
              | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 YAPNEVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 78

Qy         58 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 138

Qy        118 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 198

Qy        178 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 258

Qy        238 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 318

Qy        298 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 378

Qy        358 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 438

Qy        418 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 498
```

```
Qy        478 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 558

Qy        538 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 618

Qy        598 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 678

Qy        658 GKSTYGY 664
              |||||||
Db        679 GKSTYGY 685
```

RESULT 4
AAW61246
ID AAW61246 standard; Protein; 1007 AA.
XX
AC AAW61246;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP122 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27431.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 11; Page 92-93; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 19; Length 1007;
Best Local Similarity 100.0%; Pred. No. 8e-60;
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              |||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627
```

RESULT 5
ABP54664
ID ABP54664 standard; Protein; 1007 AA.
XX
AC ABP54664;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP122 protein sequence SEQ ID NO:216.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.

```
PA   (DOUG/) DOUGHERTY B.
PA   (FANN/) FANNON M R.
PA   (ROSE/) ROSEN C A.
XX
PI   Choi GH,  Kunsch CA,  Barash SC,  Dillon PJ,  Dougherty B,  Fannon MR;
PI   Rosen CA;
XX
DR   WPI; 2002-479261/51.
DR   N-PSDB; ABQ84899.
XX
PT   New Streptococcus pneumoniae antigens, useful for detecting
PT   Streptococcus and for preventing or attenuating disease caused by
PT   Streptococcus infection -
XX
PS   Claim 11; Page 50; 70pp; English.
XX
CC   ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC   Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC   The S. pneumoniae antigens have antibacterial activity and can be
CC   used in vaccines. The S. pneumoniae antigens can also be used to
CC   prevent or attenuate a Streptococcal infection in an animal. The
CC   polynucleotides encoding the S. pneumoniae antigens can be used to
CC   detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC   primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC   which are used in an example from the present invention.
XX
SQ   Sequence   1007 AA;

  Query Match             94.8%;  Score 4171;  DB 23;  Length 1007;
  Best Local Similarity   100.0%;  Pred. No. 8e-60;
  Matches  627;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              ||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627


RESULT 6
AAY86159
ID   AAY86159 standard; Protein; 192 AA.
XX
AC   AAY86159;
XX
DT   10-APR-2000  (first entry)
XX
DE   S. pneumoniae derived protein #368.
XX
KW   Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW   bacterial; antimicrobial; antibiotic; pathogenesis; infection.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9806734-A1.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US14436.
XX
PR   16-AUG-1996;   96US-0024022.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-159452/14.
DR   N-PSDB; AAZ96473.
XX
PT   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
PS   Claim 5; Page 617; 640pp; English.
XX
CC   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC   AAY85792-Y86182). The DNA, vectors and host cells described in the
CC   method of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
SQ   Sequence   192 AA;

  Query Match             22.2%;  Score 975.5;  DB 19;  Length 192;
  Best Local Similarity   93.7%;  Pred. No. 3.3e-08;
  Matches  163;  Conservative    1;  Mismatches    7;  Indels    3;  Gaps    2;

Qy          1 Y-P--VVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 57
              | |  ||||||| ||||| |||||||||||||| |||||||||||||||||||||||||
Db         19 YAPNEVVLADTSNSEDALSISDKEKVAENKEKHGNIHSAMETSQDFKEKKTAVIKEKEVV 78

Qy         58 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 117
              |||||||||||||||||  ||| ||||||||||||||||||||||||||||||||||||||
Db         79 SKNPVIDNNTSNEEAKIKGGNSNQSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 138

Qy        118 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQP 171
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||| |
Db        139 KAIKQLSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVHP 192
```

```
RESULT 7
AAR10559
ID   AAR10559 standard; Protein; 1962 AA.
XX
AC   AAR10559;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (A137G/K138L/T139A).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1962
FT                   /label= mature_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10413.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The mutant protease having mew cleavage specificities is
CC   obtained by replacing three amino acids.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1962 AA;

  Query Match             21.9%;  Score 962.5;  DB 12;  Length 1962;
  Best Local Similarity   26.1%;  Pred. No. 3.2e-05;
  Matches  182;  Conservative  134;  Mismatches  241;  Indels  139;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    |:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
                | :|  :  : :     : |    |  |:|  | : :  |  |: ::        |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227
              : |     |      |: :  :|           : | | |:| ||:| |  || ||: ||
Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
              :  :: | |: | ||      ::  |:|  |||  |   ||    || |  : ||||:
Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341
              |||:  | |  |       : |:|| ||::: | :|::  ||:||  | : |||||  |  :
Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSGLAGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400
              ||:  | |  :|:  : ::  :|::   ::|  :|:::||  || |:: :  |  : : :
Db        344 DVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449
              |:  |    :   | :||||:|  |    |:|        | |:  :      |: ||
Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508
              |    :| :           : ||: :|:   | :||||:  |  :: | |  :  |:
Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566
              || :::|||       ::|  |:        |   :| :|   | || |    ||
Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625
              || :       :|: | |  |: |              ||
Db        561 GVKIT-----LAMLPNQKYTEDKMS-------------DF------------------- 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
              ||:||  :| :|||::||| || || |  || ::
Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615


RESULT 8
AAR10941
ID   AAR10941 standard; Protein; 1968 AA.
XX
AC   AAR10941;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (delta137-139/ins9).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= sig_peptide
FT   Protein         188..1968
FT                   /label= mat_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10871.
XX
```

```
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29pp; English.
XX
CC   The wild-type L.lactis SK11 protease gene sequence was determined
CC   by the applicant (EP-307011).
CC   The mutant protease having mew cleavage specificities is
CC   obtained by deleting three amino acids (nine bps) and inserting
CC   9 other residues.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1968 AA;

  Query Match             21.7%;  Score 954.5;  DB 12;  Length 1968;
  Best Local Similarity   26.1%;  Pred. No. 4.4e-05;
  Matches  183;  Conservative  133;  Mismatches  241;  Indels  145;  Gaps   35;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    .|:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
                | :| :  : :     : |    |  |:|  | : :  |  |: ::        |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227
              : |    |     |: :  :|           : | | |:| ||:|  |  || ||: ||
Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
              :  :: | |: | ||      ::  |:|  |||  |   ||    ||  |  : ||||:
Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMY--SD----AG-SGFAGDETMFHAIED 335
              |||:  | |  |      : |:|| ||::: | :  ||    || :| :|  | : ||||
Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSGSAGTTGTTGSATVVSAIED 343

Qy        336 SIKHNVDVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVA 394
              | |  :||:  | |  :|:  : ::  :|::   ::|  :|:::||  || |:: :  |
Db        344 SAKIGADVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VN 401

Qy        395 NNHLKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---I 443
              : : : |:  |    :   | :||||:|  |    |:|       | |:  :
Db        402 KDYYGLQDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFT 461

Qy        444 GAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAF 502
              |: ||  |    :| :          : ||: :|:   | :||||:  |  :: | |  :
Db        462 GS-FDQKKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKY 511

Qy        503 KKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMI 560
                |:  ||  :::|||       ::|  |:       |  :| :|   | ||  |
Db        512 AQAA--GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTA 560

Qy        561 NPDKKTEVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKE 619
              ||    || :           :|: | |  |: |              ||
Db        561 HPDDSLGVKIT-----LAMLPNQKYTEDKMS-------------DF------------ 588

Qy        620 DIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
                    ||:||  :| :|||::||| || || |  || ::
Db        589 --------TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 621


RESULT 9
AAR10557
ID   AAR10557 standard; Protein; 1962 AA.
XX
AC   AAR10557;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (A137G/K138D).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1962
FT                   /label= mature_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10411.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The mutant protease A137G/K138D having mew cleavage specificities is
CC   obtained by replacing two amino acids.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1962 AA;

  Query Match             21.7%;  Score 952.5;  DB 12;  Length 1962;
  Best Local Similarity   26.0%;  Pred. No. 4.8e-05;
  Matches  181;  Conservative  134;  Mismatches  242;  Indels  139;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    |:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
                | :| :  : :     : |    |  |:|  | : :  |  |: ::        |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |
```

```
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227

Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282

Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341

Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSGDTGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400

Db        344 DVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449

Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508

Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566

Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625

Db        561 GVKIT-----LAMLPNQKYTEDKMS--------------DF------------------- 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661

Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615
```

RESULT 10
AAR10940
ID AAR10940 standard; Protein; 1974 AA.
XX
AC AAR10940;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (delta137-139/ins15).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= SIG_PEPTIDE
FT Protein 188..1974
FT /label= MAT_PROTEIN
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR WPI; 1991-038622/06.
DR N-PSDB; AAQ10870.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29pp ; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease having mew cleavage specificities is
CC obtained by deleting three amino acids (nine bps) and inserting
CC 15 other residues.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1974 AA;

Query Match 21.7%; Score 952.5; DB 12; Length 1974;
Best Local Similarity 25.8%; Pred. No. 4.8e-05;
Matches 183; Conservative 133; Mismatches 241; Indels 151; Gaps 35;

```
Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58

Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118

Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168

Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227

Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282

Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMY--SD-------AG----SGFAGDETM 329

Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSGKTGLAGKTGKTGKTGSATV 343

Qy        330 FHAIEDSIKHNVDVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSS 388

Db        344 VSAIEDSAKIGADVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSAT 403

Qy        389 SWDLVANNHLKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKY 440

Db        404 EG--VNKDYYGLQDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQL 461

Qy        441 RN---IGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTK 496

Db        462 SSHDFTGS-FDQKKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSF 511

Qy        497 DLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL 556

Db        512 DDKQKYAQAA--GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKL 560

Qy        557 --WNMINPDKKTEVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTD 613

Db        561 VDWVTAHPDDSLGVKIT-----LAMLPNQKYTEDKMS--------------DF------- 594
```

```
Qy        614 KELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
                          ||:||  :| :|||::||| || || |  || ::
Db        595 -------------TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 627


RESULT 11
AAR10558
ID   AAR10558 standard; Protein; 1962 AA.
XX
AC   AAR10558;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (A137G/K138P/T139P).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1962
FT                   /label= mature_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10412.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The mutant protease having mew cleavage specificities is
CC   obtained by replacing three amino acids.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1962 AA;

  Query Match             21.6%;  Score 951.5;  DB 12;  Length 1962;
  Best Local Similarity   26.0%;  Pred. No. 5e-05;
  Matches  181;  Conservative  133;  Mismatches  243;  Indels  139;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    |:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
                | :| :  : :     : |    |  |:|  | : :  |  |: ::         |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227
              : |    |     |: :  :|           : | | |:| ||:| |  || ||: ||
Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
              :  :: | |: | ||      ::  |:|  ||| |  ||    ||  |  : ||||:
Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341
              |||:  | |  |      : |:|| ||::: | :|:: ||  |  | : |||||  |  :
Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSGPPGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400
              ||:  | |  :|:  : ::  :|::   ::|  :|:::||  || |:: :  |  : : :
Db        344 DVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449
               |:  |    :  | :||||:|  |    |:|         | |:  :      |: ||
Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508
               |    :| :           : ||: :|:   | :||||:  |  :: | |  :  |:
Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566
              ||  :::|||        ::|  |:         |   :| :|   | ||  |    ||
Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625
               || :       :|: | |  |: |              ||
Db        561 GVKIT-----LAMLPNQKYTEDKMS-------------DF------------------- 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
                ||:||  :| :|||::||| || || |  || ::
Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615


RESULT 12
AAR10560
ID   AAR10560 standard; Protein; 1962 AA.
XX
AC   AAR10560;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (K138N).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1962
FT                   /label= mature_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
```

```
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10414.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC   The mutant protease K138N having mew cleavage specificities is
CC   obtained by carrying out single amino acid substitutions.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1962 AA;

  Query Match             21.6%;  Score 948.5;  DB 12;  Length 1962;
  Best Local Similarity   25.9%;  Pred. No. 5.6e-05;
  Matches  180;  Conservative  134;  Mismatches  243;  Indels  139;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    |:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
               | :|  :  : :     : |    |  |:|  | : :  |  |: ::        |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227
              : |    |     |: :  :|          : | | |:| ||:| |  || ||: ||
Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
               :  :: | |: | ||      ::  |:|  |||  |   ||    ||  |  : ||||:
Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341
              |||:  | |  |      : |:|| ||::: | :|::  |  :|  | : ||||| |  :
Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSANTGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400
              ||:  | |  :|:  : ::  :|::   ::|  :|:::||  || |:: :  |  : : :
Db        344 DVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449
               |:  |    :  | :||||:|  |    |:|         | |:  :       |: ||
Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508
               |    :| :          : ||: :|:   | :|||:  |  :: | |  :  |:
Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566
              ||  :::|||        ::|  |:         |   :| :|   | ||  |    ||
Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625
               || :       :|: | |  |: |              ||
Db        561 GVKIT-----LAMLPNQKYTEDKMS-------------DF------------------ 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
```

```
                ||:||  :| :|||::||| || || |  || ::
Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615


RESULT 13
AAR10561
ID   AAR10561 standard; Protein; 1962 AA.
XX
AC   AAR10561;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (N166D).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1962
FT                   /label= mature_protein
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10415.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC   The mutant protease N166D having mew cleavage specificities is
CC   obtained by carrying out single amino acid substitutions.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1962 AA;

  Query Match             21.6%;  Score 948.5;  DB 12;  Length 1962;
  Best Local Similarity   25.9%;  Pred. No. 5.6e-05;
  Matches  180;  Conservative  134;  Mismatches  243;  Indels  139;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |    |:          :|:::: |     ||       ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
               | :|  :  : :     : |    |  |:|  | : :  |  |: ::        |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
```

```
Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227

Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282

Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341

Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSANTGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSGF-TGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400

Db        344 DVLNMSLGSDSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449

Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508

Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566

Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625

Db        561 GVKIT-----LAMLPNQKYTEDKMS--------------DF------------------- 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661

Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615
```

RESULT 14
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
XX
AC AAR10563;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR WPI; 1991-038622/06.
DR N-PSDB; AAQ10417.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC The mutant protease K748T having mew cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1962 AA;

Query Match 21.6%; Score 948.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 5.6e-05;
Matches 180; Conservative 134; Mismatches 243; Indels 139; Gaps 33;

```
Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58

Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118

Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168

Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227

Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282

Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKHNV 341

Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDTSAKTGSATVVSAIEDSAKIGA 343

Qy        342 DVVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKM 400

Db        344 DVLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGL 401

Qy        401 TDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDK 449

Db        402 QDNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460

Qy        450 SKITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDK 508

Db        461 KKFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA-- 509

Qy        509 GARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKT 566

Db        510 GAAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL 560

Qy        567 EVKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPA 625

Db        561 GVKIT-----LAMLPNQKYTEDKMS--------------DF------------------- 582

Qy        626 GSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
```

```
Db        583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615


RESULT 15
AAR10562
ID   AAR10562 standard; Protein; 1959 AA.
XX
AC   AAR10562;
XX
DT   15-APR-1991  (first entry)
XX
DE   Mutant protease (delta137-139).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..187
FT                   /label= signal_peptide
FT   Protein         188..1959
FT                   /label= mature_peptide
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ,  Siezen RJ,  De Vos WM,  Kok J,  Venema G,  Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   N-PSDB; AAQ10416.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The mutant protease having mew cleavage specificities is
CC   obtained by deleting three amino acids.
CC   This mutant may then be used to prepare hybrid proteases,
CC   the fusion being between a type I and a type III protease of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s).  The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence   1959 AA;

  Query Match             21.5%;  Score 947;  DB 12;  Length 1959;
  Best Local Similarity   25.8%;  Pred. No. 5.9e-05;
  Matches  179;  Conservative  135;  Mismatches  241;  Indels  140;  Gaps   33;

Qy          1 YPV--VLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS 58
              :||  : | :: |: :|  |     |:          :|:::: |     ||        ::
Db         24 LPVGEIQAKAAISQQTLGSSLANTVTAA--------TAKQAATD----TTA-------AT 64

Qy         59 KNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEK 118
               | :|  :  : :     : |     |  |:|  | : :  |  |: ::         |  ::
Db         65 TNQAIATQLAAKGIDYNKLNKVQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQQE 123

Qy        119 AIKELSSLKNTKVLY----------TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQK 168
              : | :::   :|: :          :|  ::|| : ::   :: |:||| |: :|  |

Db        124 TNKVIAAQASVKAAVEQVTQQTAGESYGYVVNGFSTKVRVVDIPKLKQIAGVKTVTLAKV 183

Qy        169 VQPMMNHARKEIGVEEA-IDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDA 227
              : |    |     |: :  :|          : | | |:| ||:| | || ||: ||
Db        184 YYPTDAKANSMANVQAVWSNY----------KYKGEGTVVSVIDSGIDPTHKDMRLSDDK 233

Qy        228 KASMRFKKEDL-KGTDK---NYWLSDKIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
              :  :: | |: | ||      ::  |:|  |||  |   ||    ||  |  : ||||:
Db        234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTIT----DDKVD--EQHGMHV 285

Qy        283 AGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVD 342
              |||:  | |  |      : |:|| ||::: | :|:: :  :|  | : |||| |  :|
Db        286 AGIIGANGTGDDPA--KSVVGVAPEAQLLAMKVFSNSDT--SGSATVVSAIEDSAKIGAD 341

Qy        343 VVSVSSG-FTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMT 401
              |:  | |  :|:  : ::  :|::   ::|  :|:::||  || |:: :  |  : : :
Db        342 VLNMSLGSNSGNQTLEDPELAAVQNANESGTAAVISAGNSGTSGSATEG--VNKDYYGLQ 399

Qy        402 DTGNVTRTAAHEDAIAVASAKNQTVEFDKVNI--------GGESFKYRN---IGAFFDKS 450
              |: |    :   | :||||:|  |    |:|         | |:  :       |: ||
Db        400 DNEMVGSPGTSRGATTVASAENTDVITQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQK 458

Qy        451 KITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRI-YTKDLKNAFKKAMDKG 509
              |    :| :           : ||: :|:    | :||||:  |  :: | |  :  |:  |
Db        459 KFYIVKDASGN--------LSKGALADYTA-DAKGKIAIVKRGEFSFDDKQKYAQAA--G 507

Qy        510 ARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKL--WNMINPDKKTE 567
              |  :::|||        ::|  |:        |   :| :|   | ||  |    ||
Db        508 AAGLIIVNT-------DGTATPMT--SIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSLG 558

Qy        568 VKRNNKEDFKDKLE-QYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAG 626
              || :        :|: | |  |: |              ||
Db        559 VKIT-----LAMLPNQKYTEDKMS-------------DF------------------- 579

Qy        627 STSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKST 661
               ||:||  :| :|||::||| || || |  || ::
Db        580 -TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 612


Search completed: October  9, 2003, 15:51:10
Job time : 52 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:48:38 ; Search time 85 Seconds
(without alignments)
1239.935 Million cell updates/sec

Title: US-09-590-991-6
Perfect score: 4398
Sequence: 1 YPVVLADTSSSEDALNISDK.....PGKNIKSTLNVINGKSTYGY 664

Scoring table: BLOSUM30
Gapop 8.0 , Gapext 0.1

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4398	100.0	664	22	AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21	AAY81710	Streptococcus pneu
3	4381.7	99.6	2140	24	ABU01020	S. pneumoniae type
4	4171	94.8	1007	19	AAW61246	Streptococcus pneu
5	4171	94.8	1007	23	ABP54664	S. pneumoniae SP12
6	1189.3	27.0	2366	19	AAW68388	Clostridium diffic
7	1188.3	27.0	2366	17	AAR95011	C. difficile toxin
8	1186.4	27.0	3132	22	AAB51367	Japanese yam mosai
9	1175.3	26.7	6619	22	ABG23329	Novel human diagno
10	1173	26.7	5909	22	ABG23295	Novel human diagno
11	1167.8	26.6	7107	22	ABB58144	Drosophila melanog
12	1165.4	26.5	2835	23	ABB98574	Dextran saccharase
13	1165.2	26.5	31267	24	ABG74786	Human RGS11 protei
14	1164.4	26.5	4345	22	ABB66417	Drosophila melanog
15	1161.5	26.4	26926	22	AAU05396	Human titin (conne
16	1156.9	26.3	2946	22	ABG25790	Novel human diagno
17	1156.1	26.3	5464	22	ABG11810	Novel human diagno
18	1153.3	26.2	3781	22	ABG23291	Novel human diagno
19	1151.9	26.2	4643	22	ABB71609	Drosophila melanog
20	1151.8	26.2	2710	17	AAR95016	C. difficile toxin
21	1151.8	26.2	2710	19	AAW68387	Clostridium diffic
22	1151.4	26.2	3909	23	ABP43661	mRNA for hPat prot
23	1151.4	26.2	4590	23	ABP68608	Human pancreatic c
24	1151.4	26.2	4590	24	AAO26790	Human cadherin (CA
25	1151.4	26.2	4590	24	AAO26791	Human cadherin (CA
26	1151.4	26.2	4591	22	ABG22977	Novel human diagno
27	1150.9	26.2	2233	24	ABU01029	S. pneumoniae type
28	1150.9	26.2	2234	21	AAY81502	Streptococcus pneu
29	1148.8	26.1	3269	24	ABP54436	Papaya leaf-distor
30	1147.6	26.1	15281	15	AAR44929	T. niveum Cyclospo
31	1147.4	26.1	4081	22	ABB62958	Drosophila melanog
32	1147.3	26.1	3457	15	AAR62504	Large polyprotein
33	1142.5	26.0	3443	20	AAW84559	Polyprotien encode
34	1142.1	26.0	2681	24	ABJ19025	Pathogen specific
35	1140.4	25.9	4796	22	ABB58665	Drosophila melanog
36	1138.9	25.9	4555	23	AAM52106	Rat fat 3 protein
37	1134.7	25.8	4599	24	ABP56837	Human LRP1B protei
38	1133.5	25.8	4472	22	ABB60101	Drosophila melanog
39	1132.3	25.7	3457	20	AAW84560	Polyprotein encode
40	1131.7	25.7	4636	22	AAE11937	Human CG168 (or C5
41	1130.2	25.7	3542	22	AAB62142	P. falciparum FCR3
42	1130.2	25.7	7718	22	ABG11811	Novel human diagno
43	1128.8	25.7	3097	22	ABB62967	Drosophila melanog
44	1127.1	25.6	4152	22	AAG84992	Shrimp white spot
45	1126.6	25.6	5635	23	ABP60991	Novel human protei

ALIGNMENTS

RESULT 1
AAB48342
ID AAB48342 standard; Protein; 664 AA.
XX
AC AAB48342;
XX
DT 20-APR-2001 (first entry)
XX
DE S. pneumoniae Sp128 polypeptide.
XX
KW Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KW antibacterial; auditory; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO200076540-A2.
XX
PD 21-DEC-2000.
XX
PF 09-JUN-2000; 2000WO-US15925.
XX
PR 10-JUN-1999; 99US-0138453.
XX
PA (MEDI-) MED IMMUNE INC.
XX
PI Adamou JE, Choi GH;
XX
DR WPI; 2001-112197/12.
DR N-PSDB; AAC84741.
XX
PT New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT blood infections -
XX
PS Claim 4; Page 47-50; 54pp; English.
XX
CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Sp128 polypeptide.
XX
SQ Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 3.7e-73;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180

Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

RESULT 2
AAY81710
ID AAY81710 standard; Protein; 2120 AA.
XX
AC AAY81710;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID3.
XX
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006738-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB02452.
XX
PR 27-JUL-1998; 98GB-0016336.
PR 19-MAR-1999; 99US-0125329.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
DR WPI; 2000-195301/17.
DR N-PSDB; AAZ91806.
XX
PT Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections -
XX
PS Claim 2; Page 41-42; 76pp; English.
XX
CC This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments)are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis.
XX
SQ Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 3.3e-71;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
```

```
Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

```
RESULT 3
ABU01020
ID   ABU01020 standard; Protein; 2140 AA.
XX
AC   ABU01020;
XX
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #590.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media;
KW   ear infection; antiinflammatory; antibacterial; immunostimulant;
KW   auditory; respiratory; gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB02163.
XX
PR   27-MAR-2001; 2001GB-0007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX06302.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT   or ear infection -
XX
PS   Claim 1; SEQ ID No 1180; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the
CC   proteins, treating a patient by administering the protein, DNA or
CC   antibody (in a composition), a kit comprising first and second primers,
CC   which are the nucleic acid cited above or fragments between nucleotides
CC   8-100 of a sequence not defined in the specification, for amplifying a
CC   target sequence contained within a Streptococcus nucleic acid sequence,
CC   where the first primer is substantially complementary to the target
CC   sequence and the second primer is substantially complementary to the
CC   complement of the target sequence, and where the parts of the primers
CC   having substantial complementarity define the termini of the target
CC   sequence to be amplified, assay comprising contacting a test compound
CC   with the protein, and determining whether the test compound binds to the
CC   protein and a Streptococcus  pneumoniae bacterium, where one or more
CC   genes encoding the proteins has been rendered inactive. The proteins,
CC   nucleic acid molecules, antibody and compositions are useful as
CC   medicaments for treating or preventing a disease or infection due to
CC   streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC   sepsis, otitis media or ear infection. They are also useful in developing
CC   vaccines, diagnostics and antibiotics. The methods are useful for
CC   identifying immunodominant proteins. The present sequence is one of
CC   the 2469 proteins expressed by the identified coding regions from the
CC   genomic sequence.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2140 AA;
```

```
  Query Match             99.6%;  Score 4381.7;  DB 24;  Length 2140;
  Best Local Similarity   99.6%;  Pred. No. 7e-71;
  Matches  664;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    2;

Qy          1 Y-P--VVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 57
              | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 YAPNEVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 78

Qy         58 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEPKDKESGE 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEPKDKESGE 138

Qy        118 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 198

Qy        178 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 258

Qy        238 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 318

Qy        298 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 378

Qy        358 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 438

Qy        418 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 498
```

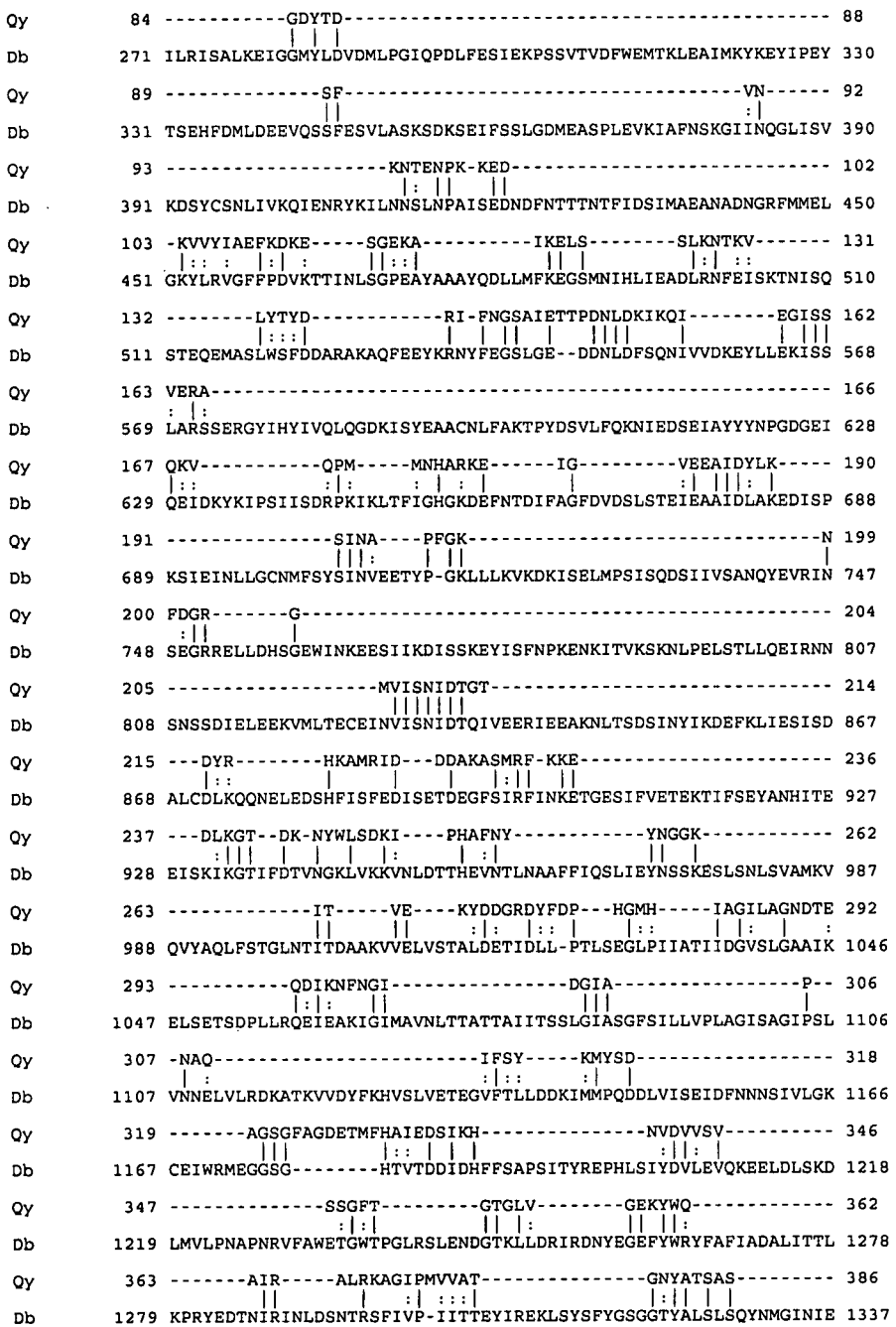
```
Qy        478 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 558

Qy        538 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 618

Qy        598 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 678

Qy        658 GKSTYGY 664
              |||||||
Db        679 GKSTYGY 685
```

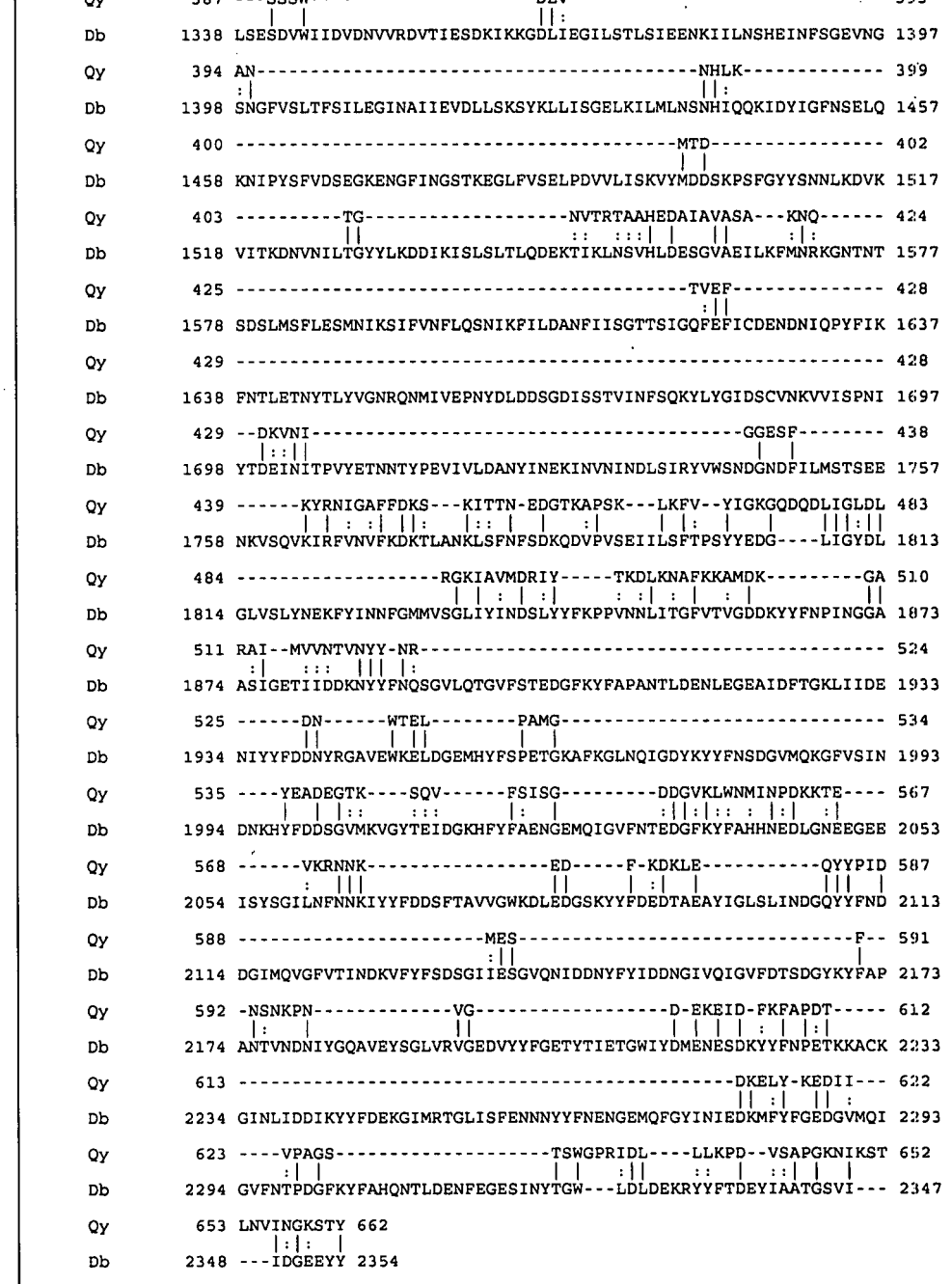
```
RESULT 4
AAW61246
ID   AAW61246 standard; Protein; 1007 AA.
XX
AC   AAW61246;
XX
DT   02-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae SP122 protein.
XX
KW   Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW   detection; pneumonia; otitis media; meningitis.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818930-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19422.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Choi GH,  Hromockyj A,  Johnson LS,  Kunsch CA;
XX
DR   WPI; 1998-272224/24.
DR   N-PSDB; AAV27431.
XX
PT   Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT   pneumoniae - or their epitope-containing fragments, useful in
PT   protective or therapeutic vaccines, and for diagnosis
XX
PS   Claim 11; Page 92-93; 118pp; English.
XX
CC   The present sequence represents a protein from Streptococcus pneumoniae.
CC   The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC   can be useful in vaccines for inducing protective antibodies against
CC   Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC   pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC   are used to detect Streptococcus infection (by usual hybridisation or
CC   amplification methods), also for isolating Streptococcus genes or their
CC   allelic variants. The protein can be used similarly to detect specific
CC   antibodies in standard immunoassays, especially for diagnosing or
CC   monitoring infections. Antibodies which bind the protein are used to
CC   detect corresponding antigens, to purify the protein and for passive
CC   immunisation (optionally coupled to a toxin). Vaccines are administered,
CC   e.g. by injection, orally or through the skin, typically at 0.01-1000
CC   (especially 10-300) mu g/ml per dose.
XX
SQ   Sequence   1007 AA;

  Query Match            94.8%;  Score 4171;  DB 19;  Length 1007;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-68;
  Matches  627;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              |||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627
```

```
RESULT 5
ABP54664
ID   ABP54664 standard; Protein; 1007 AA.
XX
AC   ABP54664;
XX
DT   04-SEP-2002  (first entry)
XX
DE   S. pneumoniae SP122 protein sequence SEQ ID NO:216.
XX
KW   Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW   antibacterial; Streptococcal infection; detection.
XX
OS   Streptococcus pneumoniae.
XX
PN   US2002061545-A1.
XX
PD   23-MAY-2002.
XX
PF   22-JAN-2001; 2001US-0765272.
XX
PR   30-OCT-1997;   97US-0961083.
XX
PA   (CHOI/) CHOI G H.
PA   (KUNS/) KUNSCH C A.
PA   (BARA/) BARASH S C.
PA   (DILL/) DILLON P J.
```

```
PA   (DOUG/) DOUGHERTY B.
PA   (FANN/) FANNON M R.
PA   (ROSE/) ROSEN C A.
XX
PI   Choi GH,  Kunsch CA,  Barash SC,  Dillon PJ,  Dougherty B,  Fannon MR;
PI   Rosen CA;
XX
DR   WPI; 2002-479261/51.
DR   N-PSDB; ABQ84899.
XX
PT   New Streptococcus pneumoniae antigens, useful for detecting
PT   Streptococcus and for preventing or attenuating disease caused by
PT   Streptococcus infection -
XX
PS   Claim 11; Page 50; 70pp; English.
XX
CC   ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC   Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC   The S. pneumoniae antigens have antibacterial activity and can be
CC   used in vaccines. The S. pneumoniae antigens can also be used to
CC   prevent or attenuate a Streptococcal infection in an animal. The
CC   polynucleotides encoding the S. pneumoniae antigens can be used to
CC   detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC   primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC   which are used in an example from the present invention.
XX
SQ   Sequence   1007 AA;

  Query Match             94.8%;  Score 4171;  DB 23;  Length 1007;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-68;
  Matches  627;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              ||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627


RESULT 6
AAW68388
ID   AAW68388 standard; Protein; 2366 AA.
XX
AC   AAW68388;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium difficile toxin B.
XX
KW   Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
KW   pseudomembranous enterocolitis.
XX
OS   Clostridium difficile.
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US15394.
XX
PR   28-AUG-1996;   96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS,  Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30561.
XX
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
XX
PS   Example 18; Page 241-249; 428pp; English.
XX
CC   This is the amino acid sequence of Clostridium difficile toxin B,
CC   deduced from the coding region (see AAV30561) of the toxin B gene.
CC   Fragments of the toxin B gene have been cloned into various
CC   prokaryotic expression systems, and assessed for the ability to
CC   express recombinant toxin B protein in E. coli.  It would be
CC   advantageous to use simple and inexpensive prokaryotic expression
CC   systems to produce and purify high levels of recombinant toxin B
CC   for immunisation purposes.  The invention specifically relates to
CC   recombinant proteins derived from Clostridium botulinum toxins
CC   (see AAW68389-400) and their use as immunogens for the production of
CC   vaccines and antitoxins.
XX
SQ   Sequence   2366 AA;

  Query Match             27.0%;  Score 1189.3;  DB 19;  Length 2366;
  Best Local Similarity   12.4%;  Pred. No. 3.8e-09;
  Matches  292;  Conservative  143;  Mismatches  198;  Indels 1717;  Gaps  119;

Qy          1 YP------VV-----LADTSSSEDALNISDKEKVAENK---------------------- 27
              |:      ||     | |  |  | : |   :|   ||
Db         34 YHNMSENTVVEKYLKLKDINSLTDIY-IDTYKKSGRNKALKKFKEYLVTEVLELKNNNLT 92

Qy         28 --EKHENIH-------------------------------------------------SA 36
                ||  |:|                                                 :
Db         93 PVEK--NLHFVWIGGQINDTAINYINQWKDVNSDYNVNVFYDSNAFLINTLKKTVVESAI 150

Qy         37 METSQDFKE--------------KKTAVI--KEKEVV-------SKNP--VIDNNT---- 67
              :| : |:|               |: ::|  |:| ::         :||  :||: :
Db        151 NDTLESFRENLNDPRFDYNKFFRKRMQIIYDKQKNFINYYKAQREENPELIIDDIVKTYL 210

Qy         68 SNEEAK-IKEENS------NK-SQ---------------------------------- 83
              ||| :| | | |:      || :|
Db        211 SNEYSKEIDELNTYIEESLNKITQNSGNDVRNFEEFKNGESFNLYEQELVERWNLAAASD 270
```





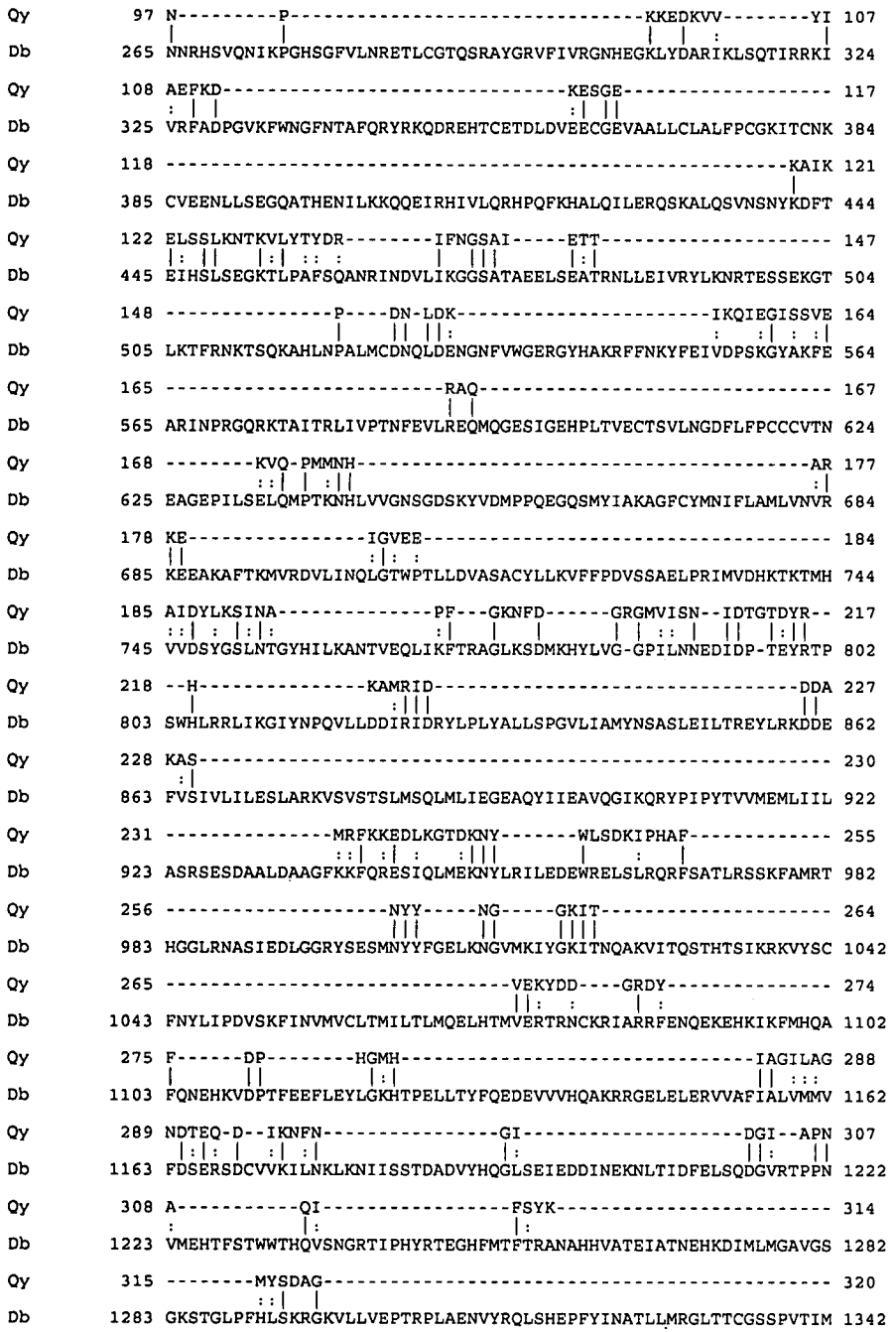
RESULT 7
AAR95011
ID AAR95011 standard; Protein; 2366 AA.
XX
AC AAR95011;
XX
DT 08-JUL-1996 (first entry)
XX
DE C. difficile toxin B.
XX
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW diarrhoea; therapy; diagnosis; vaccine.
XX
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US13737.
XX
PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29247.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin -
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX
PS Claim 36; Page 313-323; 434pp; English.
XX
CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
CC product of the toxin B gene (AAT29247), is a cytotoxin associated
CC with diarrhoetic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
CC AAR75371-72 and AAR95018), pref. expressed as fusions to polyhistidine
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
XX
SQ Sequence 2366 AA;

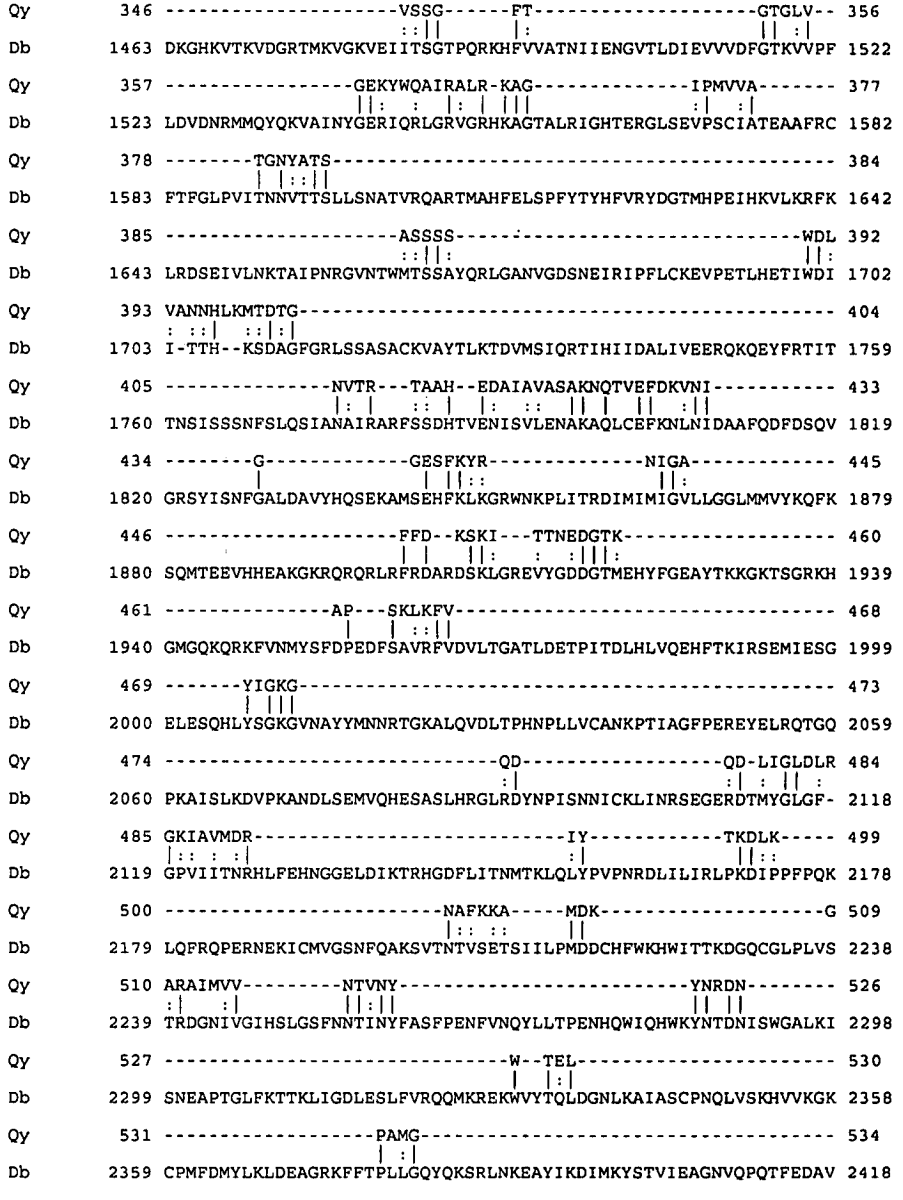
Query Match 27.0%; Score 1188.3; DB 17; Length 2366;
Best Local Similarity 12.4%; Pred. No. 4e-09;
Matches 292; Conservative 142; Mismatches 199; Indels 1717; Gaps 119;

```
Qy          1 YP------VV-----LADTSSSEDALNISDKEKVAENK---------------------- 27
              |:      ||     | | |  | : |   :|   ||
Db         34 YHNMSENTVVEKYLKLKDINSLTDIY-IDTYKKSGRNKALKKFKEYLVTEVLELKNNNLT 92

Qy         28 --EKHENIH-------------------------------------------------SA 36
                ||  |:|                                                 :
Db         93 PVEK--NLHFVWIGGQINDTAINYINQWKDVNSDYNVNVFYDSNAFLINTLKKTVVESAI 150

Qy         37 METSQDFKE--------------KKTAVI--KEKEVV-------SKNP--VIDNNT---- 67
                :| : |:|              |:  :|  |:| ::        :||  :||: :
Db        151 NDTLESFRENLNDPRFDYNKFFRKRMEIIYDKQKNFINYYKAQREENPELIIDDIVKTYL 210

Qy         68 SNEEAK-IKEENS------NK-SQ---------------------------------- 83
              ||| :| | | |:      || :|
Db        211 SNEYSKEIDELNTYIEESLNKITQNSGNDVRNFEEFKNGESFNLYEQELVERWNLAAASD 270

Qy         84 -----------GDYTD------------------------------------------- 88
                                                                  | | |
Db        271 ILRISALKEIGGMYLDVDMLPGIQPDLFESIEKPSSVTVDFWEMTKLEAIMKYKEYIPEY 330

Qy         89 --------------SF-----------------------------------VN------ 92
                          ||                                    :|
Db        331 TSEHFDMLDEEVQSSFESVLASKSDKSEIFSSLGDMEASPLEVKIAFNSKGIINQGLISV 390

Qy         93 --------------------KNTENPK-KED---------------------------- 102
                                 |: ||   ||
Db        391 KDSYCSNLIVKQIENRYKILNNSLNPAISEDNDFNTTTNTFIDSIMAEANADNGRFMMEL 450

Qy        103 -KVVYIAEFKDKE-----SGEKA----------IKELS--------SLKNTKV------- 131
               |:: :  |:| :     ||::|          || |         |:| ::
Db        451 GKYLRVGFFPDVKTTINLSGPEAYAAAYQDLLMFKEGSMNIHLIEADLRNFEISKTNISQ 510

Qy        132 --------LYTYD-----------RI-FNGSAIETTPDNLDKIKQI--------EGISS 162
                     |:::|            |  | ||  |   ||||    |         | |||
Db        511 STEQEMASLWSFDDARAKAQFEEYKRNYFEGSLGE--DDNLDFSQNIVVDKEYLLEKISS 568

Qy        163 VERA------------------------------------------------------ 166
              : |:
Db        569 LARSSERGYIHYIVQLQGDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYYNPGDGEI 628

Qy        167 QKV-----------QPM-----MNHARKE------IG---------VEEAIDYLK----- 190
              |::           :|:     : | : |       |         :| |||: |
Db        629 QEIDKYKIPSIISDRPKIKLTFIGHGKDEFNTDIFAGFDVDSLSTEIEAAIDLAKEDISP 688

Qy        191 --------------SINA----PFGK-------------------------------N 199
                           |||:    | ||                               |
Db        689 KSIEINLLGCNMFSYSINVEETYP-GKLLLKVKDKISELMPSISQDSIIVSANQYEVRIN 747

Qy        200 FDGR-------G---------------------------------------------- 204
              :||       |
Db        748 SEGRRELLDHSGEWINKEESIIKDISSKEYISFNPKENKITVKSKNLPELSTLLQEIRNN 807

Qy        205 -------------------MVISNIDTGT----------------------------- 214
                                |||||||
Db        808 SNSSDIELEEKVMLTECEINVISNIDTQIVEERIEEAKNLTSDSINYIKDEFKLIESISD 867

Qy        215 ---DYR--------HKAMRID---DDAKASMRF-KKE---------------------- 236
                 |::        |     |    |    |:||  ||
Db        868 ALCDLKQQNELEDSHFISFEDISETDEGFSIRFINKETGESIFVETEKTIFSEYANHITE 927

Qy        237 ---DLKGT--DK-NYWLSDKI----PHAFNY------------YNGGK------------ 262
                  :|||  |  |  |  |:     | :|             ||  |
Db        928 EISKIKGTIFDTVNGKLVKKVNLDTTHEVNTLNAAFFIQSLIEYNSSKESLSNLSVAMKV 987

Qy        263 -------------IT-----VE----KYDDGRDYFDP---HGMH-----IAGILAGNDTE 292
                          ||     ||     :|:  |:: |    |::     | |:  |   :
Db        988 QVYAQLFSTGLNTITDAAKVVELVSTALDETIDLL-PTLSEGLPIIATIIDGVSLGAAIK 1046

Qy        293 -----------QDIKNFNGI---------------DGIA----------------P-- 306
                        |:|:   ||               |||                |
Db       1047 ELSETSDPLLRQEIEAKIGIMAVNLTTATTAIITSSLGIASGFSILLLPLAGISAGIPSL 1106

Qy        307 -NAQ-------------------------IFSY-----KMYSD---------------- 318
               | :                       :|::     :|  |
Db       1107 VNNELVLRDKATKVVDYFKHVSLVETEGVFTLLDDKIMMPQDDLVISEIDFNNNSIVLGK 1166

Qy        319 -------AGSGFAGDETMFHAIEDSIKH--------------NVDVVSV---------- 346
                    |||         |:: | | |                 :||: |
Db       1167 CEIWRMEGGSG--------HTVTDDIDHFFSAPSITYREPHLSIYDVLEVQKEELDLSKD 1218

Qy        347 --------------SSGFT---------GTGLV--------GEKYWQ------------ 362
                           :|:|          || |:         || ||:
Db       1219 LMVLPNAPNRVFAWETGWTPGLRSLENDGTKLLDRIRDNYEGEFYWRYFAFIADALITTL 1278

Qy        363 -------AIR-----ALRKAGIPMVVAT--------------GNYATSAS--------- 386
                    ||      |   :| :::|                 |:|| | |
Db       1279 KPRYEDTNIRINLDSNTRSFIVP-IITTEYIREKLSYSFYGSGGTYALSLSQYNMGINIE 1337

Qy        387 ---SSSW--------------------DLV---------------------------- 393
                  | |                    ||:
```

```
Db       1338 LSESDVWIIDVDNVVRDVTIESDKIKKGDLIEGILSTLSIEENKIILNSHEINFSGEVNG 1397

Qy        394 AN-----------------------------------------NHLK------------ 399
              :|                                         ||:
Db       1398 SNGFVSLTFSILEGINAIIEVDLLSKSYKLLISGELKILMLNSNHIQQKIDYIGFNSELQ 1457

Qy        400 ----------------------------------------MTD---------------- 402
                                                       | |
Db       1458 KNIPYSFVDSEGKENGFINGSTKEGLFVSELPDVVLISKVYMDDSKPSFGYYSNNLKDVK 1517

Qy        403 ----------TG------------------NVTRTAAHEDAIAVASA---KNQ------ 424
                       ||                  ::  :::| |   ||     :|:
Db       1518 VITKDNVNILTGYYLKDDIKISLSLTLQDEKTIKLNSVHLDESGVAEILKFMNRKGNTNT 1577

Qy        425 ---------------------------------------TVEF-------------- 428
                                                      :||
Db       1578 SDSLMSFLESMNIKSIFVNFLQSNIKFILDANFIISGTTSIGQFEFICDENDNIQPYFIK 1637

Qy        429 ------------------------------------------------------------ 428

Db       1638 FNTLETNYTLYVGNRQNMIVEPNYDLDDSGDISSTVINFSQKYLYGIDSCVNKVVISPNI 1697

Qy        429 --DKVNI-----------------------------------------GGESF-------- 438
               |::||                                         |  |
Db       1698 YTDEINITPVYETNNTYPEVIVLDANYINEKINVNINDLSIRYVWSNDGNDFILMSTSEE 1757

Qy        439 ------KYRNIGAFFDKS---KITTN-EDGTKAPSK---LKFV--YIGKGQDQDLIGLDL 483
                   | | : :| ||:   |:: |  |   :|     | |:  |   |    |||:||
Db       1758 NKVSQVKIRFVNVFKDKTLANKLSFNFSDKQDVPVSEIILSFTPSYYEDG----LIGYDL 1813

Qy        484 -------------------RGKIAVMDRIY-----TKDLKNAFKKAMDK---------GA 510
                              | | : | :|     : :| : |  : |           ||
Db       1814 GLVSLYNEKFYINNFGMMVSGLIYINDSLYYFKPPVNNLITGFVTVGDDKYYFNPINGGA 1873

Qy        511 RAI--MVVNTVNYY-NR------------------------------------------- 524
               :|   :::  ||| |:
Db       1874 ASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFTGKLIIDE 1933

Qy        525 ------DN------WTEL--------PAMG----------------------------- 534
                   ||      | ||        |  |
Db       1934 NIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQIGDYKYYFNSDGVMQKGFVSIN 1993

Qy        535 ----YEADEGTK----SQV------FSISG---------DDGVKLWNMINPDKKTE---- 567
                   | | |::    :::      |:  |         :||:|:: : |:|  :|
Db       1994 DNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEE 2053

Qy        568 ------VKRNNK----------------ED-----F-KDKLE----------QYYPID 587
                   :  |||                ||     | :|  |          |||  |
Db       2054 ISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFND 2113

Qy        588 ----------------------MES-----------------------------F-- 591
                               :||                               |
Db       2114 DGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAP 2173

Qy        592 -NSNKPN------------VG----------------D-EKEID-FKFAPDT----- 612
               |:   |            ||                | | | | : | |:|
Db       2174 ANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACK 2233

Qy        613 ------------------------------------------DKELY-KEDII--- 622
                                                       || :|  || :
Db       2234 GINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQI 2293

Qy        623 ----VPAGS-------------------TSWGPRIDL----LLKPD--VSAPGKNIKST 652
                   :| |                  | |   :||    ::  |  ::| |  |
Db       2294 GVFNTPDGFKYFAHQNTLDENFEGESINYTGW---LDLDEKRYYFTDEYIAATGSVI--- 2347

Qy        653 LNVINGKSTY 662
                |:|:  |
Db       2348 ---IDGEEYY 2354


RESULT 8
AAB51367
```

```
ID   AAB51367 standard; Protein; 3132 AA.
XX
AC   AAB51367;
XX
DT   10-APR-2001  (first entry)
XX
DE   Japanese yam mosaic virus JYMV-M protein sequence SEQ ID NO:1.
XX
KW   Japanese yam mosaic virus; JYMV; Dioscorea; resistance; JYMV-M;
KW   Dioscoreaceae mosaic virus.
XX
OS   Japanese yam mosaic virus.
XX
PN   JP2000300270-A.
XX
PD   31-OCT-2000.
XX
PF   21-APR-1999;   99JP-0113671.
XX
PR   21-APR-1999;   99JP-0113671.
XX
PA   (AICH-) AICHI KEN PREFECTURE.
XX
DR   WPI; 2001-141353/15.
DR   N-PSDB; AAF26998.
XX
PT   Weakly toxic strain of Dioscoreaceae mosaic virus used in the genetic
PT   diagnosis of the mosaic virus infection and for the management of crops
PT   resistant to the virus -
XX
PS   Claim 1; Page 6-20; 22pp; Japanese.
XX
CC   The present sequence represents the Japanese yam mosaic virus (JYMV)
CC   JYMV-M protein sequence. JYMV is a weakly toxic strain of Dioscoreaceae
CC   mosaic virus. The present invention also describes: (1) a weakly toxic
CC   strain of Dioscoreaceae mosaic virus having a nucleic acid sequence
CC   which shows interference against the infection of a strongly toxic
CC   strain of Dioscoreaceae mosaic virus; (2) a Dioscoreaceae mosaic
CC   virus-resistant crop which is a plant of Dioscorea japonica, Dioscorea
CC   opposita or Dioscorea alata and is infected artificially by a weakly
CC   toxic strain of Dioscoreaceae mosaic virus; (3) a method for the
CC   cultivation of a Dioscoreaceae mosaic virus-resistant crop in which the
CC   weakly toxic strain of Dioscoreaceae mosaic virus is infected to the
CC   tubercle or the bulbil of a plant of Dioscorea japonica, Dioscorea
CC   opposita or Dioscorea alata and the tubercle or bulbil is grown; and
CC   (4) a method for genetic diagnosis of a Dioscoreaceae mosaic virus in
CC   which a DNA fragment amplified by using reverse transcriptase-polymerase
CC   chain reaction (RT-PCR) primers atagcttacacacccctcac (AAF26999) and
CC   caccdtacattgcwgaragyg (AAF32201) is treated with a restriction enzyme
CC   to distinguish the weakly toxic strain of Dioscoreaceae mosaic virus
CC   from the strongly toxic strain of Dioscoreaceae mosaic virus.
XX
SQ   Sequence   3132 AA;

  Query Match           27.0%;  Score 1186.4;  DB 22;  Length 3132;
  Best Local Similarity   9.9%;  Pred. No. 1.3e-08;
  Matches  302;  Conservative  170;  Mismatches  180;  Indels 2386;  Gaps  132;

Qy          1 YP----------------------VVLADTSSSEDALNISDKEKV----AE-NK------ 27
              :|                     |: : :: ::  :|: :|  :    || ||
Db         26 LPPVFIGIPATSCKTIATDKNDEHVIQSKSTEAQKKMNVNNKAFIHKMHAEYNKQCAVLD 85

Qy         28 --EKHEN--------------------IHS-AMETSQDFKEKKTAVIKEKEVV-----S 58
                |:  |                    :|   || ::  | |:   |  |:::     |
Db         86 KWEEEYNTTRSTNPSSYKVVLTKQQKSAVHRYNMEKARK-KLKEQEDILNKCIIGPWGPS 144

Qy         59 KN-----------------------------------PV----------IDNNTS----- 68
              :|                                   ||          : | |:
Db        145 MNAFSIASGPLPSAMEAELKWPLHQTRSQRIKGTPKQPVTLGRGEFAKFVKNLTTLMTQK 204

Qy         69 ----------------NEEAKIK-------EENSNKSQGDYTDSF---------VNKNTE 96
                             |  |:          ||  ||::  : | |         :::::
Db        205 SLLLELCGKHVHRVCVRREHKKVYLKINTKHEEGFNKARDVVMDNFTQRLLELMITRTSG 264
```



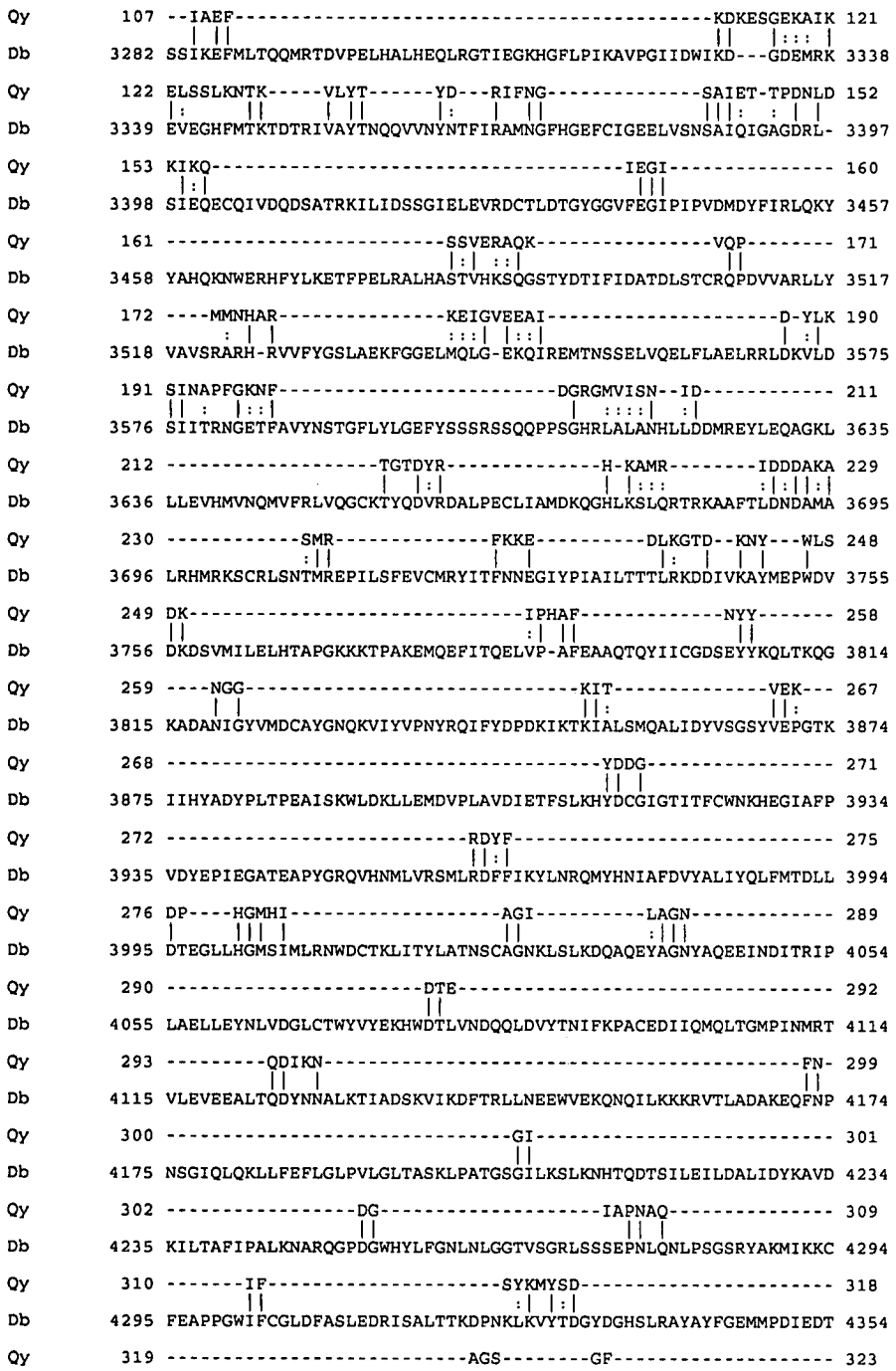


```
Db       2419 RLLIQDLSELGFETCQYITDEDVIFNALNMKSAVGALYGGKKKDYFKDFTQEMKETILKQ 2478

Qy        549 ------SGDDGVKLWN-----MINPDKKTEVKRNNK------------------EDFK- 577

Db       2479 SCARLYTGKMG--LWNGSLKAELRPLEKV---QANKTRTFTAAPLDTLLGGKACVDDFNN 2533

Qy        578 ------------------------------------------------------------ 577

Db       2534 QFYELNIKGPWSVGMTKFYGGWNELLTKLPDGWIHCDADGSQFDSSLSPYLINAVLNIRL 2593

Qy        578 ------DKLEQ----------YYPI---D---MESF---NSNKP-----NV--------- 598

Db       2594 HFMETWDIGEQMLRNLYTEIVYTPIATPDGTIVKKFKGNNSGQPSTVVDNTLMVLLALKY 2653

Qy        599 -----GDEKE-----------------ID-------------FK-------FAPDT-DK 614

Db       2654 SLLKDGVEAEKHKQVIKYFVNGDDLLISIDPAYEGLLDTMQGNFKELGLKYDFNSRTRDK 2713

Qy        615 -ELYK--------EDIIVP---------------------------AGSTSWG------ 631

Db       2714 GELWFMSHQGKRVEDIWIPKLEQERIVSILEWDRSKEPGNRMEAICAAMIESWGHQELTH 2773

Qy        632 ------------------------PRI---------------------------DLLLK 639

Db       2774 QIRRFYAWLIGQAPYSGLAEIGKAPYIAESALRKLYLDKDADQSAIEVYLRAIFEDYTTE 2833

Qy        640 P----------D--VSAP----------------------------GK------- 647

Db       2834 PEDLFVYHQSGDDTLDAGTSAPSKARKQESASSTQGIAPTMEGFEPADDPTGKSKQQTVI 2893

Qy        648 NIKSTLNV---------------------------------------------------- 655

Db       2894 NLEKDVNVGTVGTFAVPRLKGLATKMSMPRVRGKAAMNLDHLLVYNPEQVDLANTRATRK 2953

Qy        656 ----------------------------------------ING----------------- 658

Db       2954 QFDTWYDGVKRDYELDDSSMQIILNGLMVWCIENGTSPNINGMWVMMDGEEQIEYPIKPL 3013

Qy        659 ----KST---------------------------YGY 664

Db       3014 IDHAKPTFRQIMAHFSYVAEAYIEKRNQEKAYMPRYGL 3051
```

```
RESULT 9
ABG23329
ID   ABG23329 standard; Protein; 6619 AA.
XX
AC   ABG23329;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23320.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87516.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 53688; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   6619 AA;
```

Query Match 26.7%; Score 1175.3; DB 22; Length 6619;
Best Local Similarity 9.4%; Pred. No. 3.8e-07;
Matches 345; Conservative 151; Mismatches 142; Indels 3047; Gaps 153;

```
Qy          2 PV--------------VLAD-------TSSSEDALNISDK-------------------E 21

Db       2746 PVKATDVKPIKKTKMEILKDLWNQFLKAETSNEAITLADEIIAAKGKCKKSWASLGFSGD 2805

Qy         22 KVAE------NKEK---------------------------------------------- 29

Db       2806 MVQKLIDARANKPKPEKAVKQDKTVAPKKTTVTRSRNADVIRSYLNDKGMSDYCKAHAIH 2865

Qy         30 ----------------------------HENIHSAME----------------------- 38

Db       2866 DVKRAAKKSYAALGLTPEECSQIDKLKHHLKIHTMAKVYRANRKAHDDDIIRMNAVGLSL 2925

Qy         39 --------------------------TSQDFKE------------------------- 45

Db       2926 GTIAKTLGVHPTTVTLRLRSLGIEPADTRRTFMENVLAPLPTHVADWLSEQVGPKYEIRQ 2985

Qy         46 ---------------KKTA----------VIK--------------------------- 52

Db       2986 YVRDLLLEAYNNRHNHKETAHERFLKRYVEVVEMLEELAPNSEEAYDTLHNAIAALTELS 3045

Qy         53 ------EKEVVSKN--------------------------P-----VIDNNTS----- 68

Db       3046 ELMKSSDKEVYAPNSNRLAILDALADQIVTATGIGTFLGMNVPGALTEVNRSNYSKFEDG 3105

Qy         69 ----NEEAKIKEE------------------------------------------------ 77

Db       3106 EPVFNENKKVMKEAAAEGFFDFLLDPNASELRISGPGGTGKTFTMAHMIDEIMPRYHETC 3165

Qy         78 ---------------NSNKS-----------------QG-----------DYTDS-- 89

Db       3166 SLMGIPALYNEVIMTATTNKAAEVLARATGRPTSTYHSFQGLTVRNNLKTGEADLIPSKS 3225

Qy         90 ----------------------FVNKNTENPKKEDKVVY-------------------- 106

Db       3226 FSIKKNKVIFIDEASMIDRKLKKFINEGTH----QSKIVYLGDKCQLLPVKETSSPVYND 3281
```

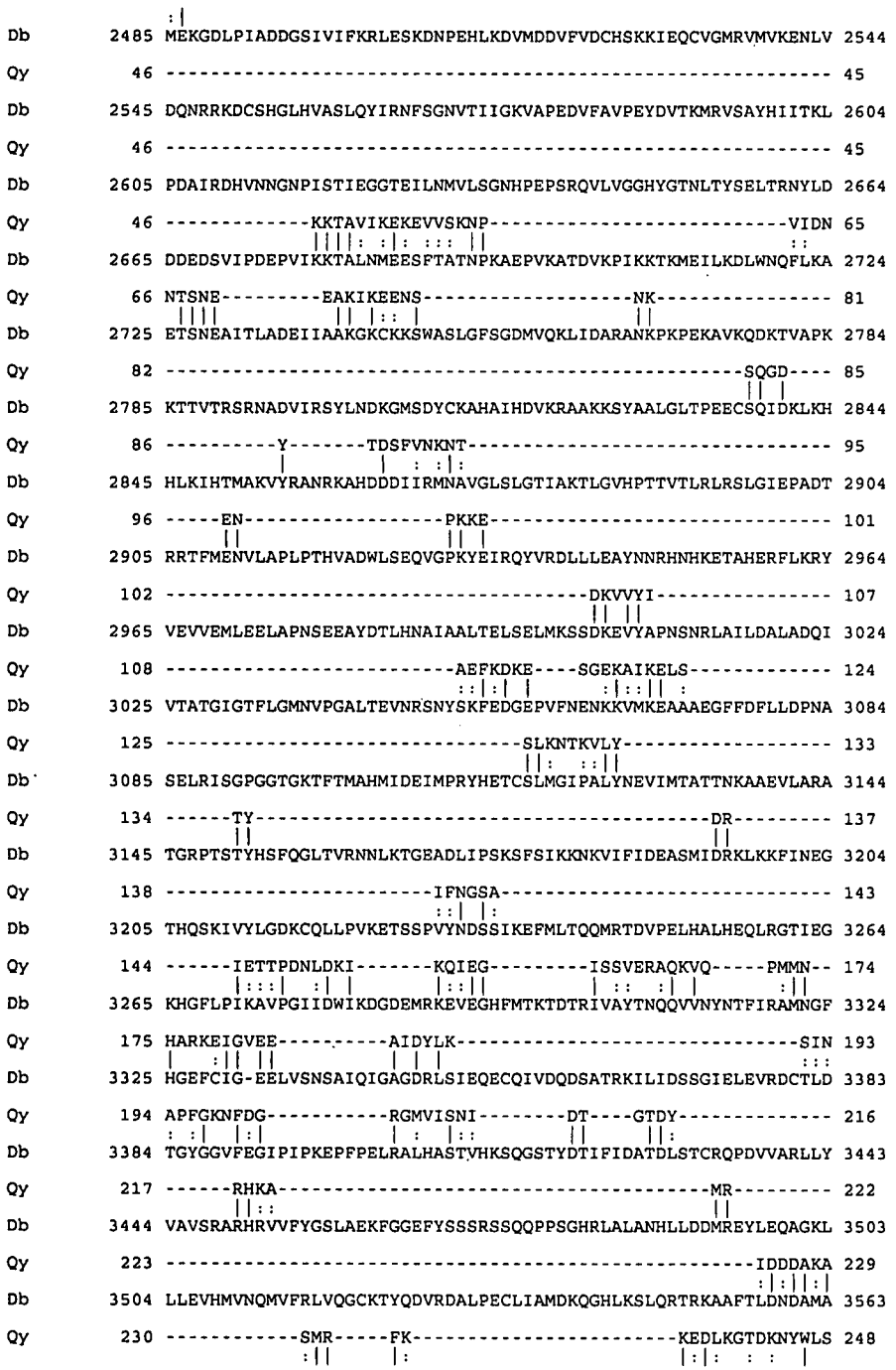

Qy 107 --IAEF---KDKESGEKAIK 121
Db 3282 SSIKEFMLTQQMRTDVPELHALHEQLRGTIEGKHGFLPIKAVPGIIDWIKD---GDEMRK 3338
Qy 122 ELSSLKNTK-----VLYT------YD---RIFNG-------------SAIET-TPDNLD 152
Db 3339 EVEGHFMTKTDTRIVAYTNQQVVNYNTFIRAMNGFHGEFCIGEELVSNSAIQIGAGDRL- 3397
Qy 153 KIKQ---IEGI--------------- 160
Db 3398 SIEQECQIVDQDSATRKILIDSSGIELEVRDCTLDTGYGGVFEGIPIPVDMDYFIRLQKY 3457
Qy 161 -------------------------SSVERAQK----------------VQP--------- 171
Db 3458 YAHQKNWERHFYLKETFPELRALHASTVHKSQGSTYDTIFIDATDLSTCRQPDVVARLLY 3517
Qy 172 ----MMNHAR---------------KEIGVEEAI--------------------D-YLK 190
Db 3518 VAVSRARH-RVVFYGSLAEKFGGELMQLG-EKQIREMTNSSELVQELFLAELRRLDKVLD 3575
Qy 191 SINAPFGKNF-------------------------DGRGMVISN--ID------------ 211
Db 3576 SIITRNGETFAVYNSTGFLYLGEFYSSSRSSQQPPSGHRLALANHLLDDMREYLEQAGKL 3635
Qy 212 -------------------TGTDYR--------------H-KAMR--------IDDDAKA 229
Db 3636 LLEVHMVNQMVFRLVQGCKTYQDVRDALPECLIAMDKQGHLKSLQRTRKAAFTLDNDAMA 3695
Qy 230 ------------SMR---------------FKKE----------DLKGTD--KNY---WLS 248
Db 3696 LRHMRKSCRLSNTMREPILSFEVCMRYITFNNEGIYPIAILTTTLRKDDIVKAYMEPWDV 3755
Qy 249 DK------------------------------IPHAF-------------NYY------- 258
Db 3756 DKDSVMILELHTAPGKKKTPAKEMQEFITQELVP-AFEAAQTQYIICGDSEYYKQLTKQG 3814
Qy 259 ----NGG-------------------------------KIT---------------VEK--- 267
Db 3815 KADANIGYVMDCAYGNQKVIYVPNYRQIFYDPDKIKTKIALSMQALIDYVSGSYVEPGTK 3874
Qy 268 --YDDG----------------- 271
Db 3875 IIHYADYPLTPEAISKWLDKLLEMDVPLAVDIETFSLKHYDCGIGTITFCWNKHEGIAFP 3934
Qy 272 ---------------------------RDYF----------------------------- 275
Db 3935 VDYEPIEGATEAPYGRQVHNMLVRSMLRDFFIKYLNRQMYHNIAFDVYALIYQLFMTDLL 3994
Qy 276 DP----HGMHI-------------------AGI----------LAGN------------- 289
Db 3995 DTEGLLHGMSIMLRNWDCTKLITYLATNSCAGNKLSLKDQAQEYAGNYAQEEINDITRIP 4054
Qy 290 ----------------------DTE---------------------------------- 292
Db 4055 LAELLEYNLVDGLCTWYVYEKHWDTLVNDQQLDVYTNIFKPACEDIIQMQLTGMPINMRT 4114
Qy 293 ---------QDIKN---FN- 299
Db 4115 VLEVEEALTQDYNNALKTIADSKVIKDFTRLLNEEWVEKQNQILKKKRVTLADAKEQFNP 4174
Qy 300 -------------------------------GI--------------------------- 301
Db 4175 NSGIQLQKLLFEFLGLPVLGLTASKLPATGSGILKSLKNHTQDTSILEILDALIDYKAVD 4234
Qy 302 -----------------DG-------------------IAPNAQ--------------- 309
Db 4235 KILTAFIPALKNARQGPDGWHYLFGNLNLGGTVSGRLSSSEPNLQNLPSGSRYAKMIKKC 4294
Qy 310 -------IF--------------------SYKMYSD---------------------- 318
Db 4295 FEAPPGWIFCGLDFASLEDRISALTTKDPNKLKVYTDGYDGHSLRAYAYFGEMMPDIEDT 4354
Qy 319 --------------------------AGS--------GF-------------------- 323
Db 4355 VESVNSIQEKYKAYRQDSKAPTFALTYQGTYITLMKNCGFPEQKARMVEERYHTLYKVSD 4414
Qy 324 -- 323
Db 4415 DWVQAKLDQAAKDGYVTVAFGLRVRTPLLQQVIRGTSKTPYEAEAEGRTAGNALGQSWCL 4474
Qy 324 ----AGDE-----------------TMFHA-----IED---------------------- 335
Db 4475 LNNRAGSEFMRKVRNSKHRLDIRPSTHIHDAQYFLIRDDMDVVIYTNTHLVKAVQWQDHP 4534
Qy 336 SIKH--- 339
Db 4535 DIAHPDVHLDGQEGAVIVQNTMLLTEEKQVTFRDLGRAQSALFQNLNDRFGQVVDVRDIV 4594
Qy 340 --NV--------------------------------DVV---------------------- 344
Db 4595 FLNVSYLGHMSQPDFQQNLVTNPQRMEQPPYQAECEDIIQALNMTFDEGCLFKALWRNAA 4654
Qy 345 --SVSSGF 350
Db 4655 ARLGNGKPGNNAVYDAEKMVHYANRIFIKESGGQKKEAEEPVSAALKKGEEIHMSLSELF 4714
Qy 351 T-- 351
Db 4715 TLLTGEAAEGEVMSLAGWHTHVPNRLGVMPIAVAMSKAEQVDVVRADGTTITHVFPHEID 4774
Qy 352 -- 351
Db 4775 WSDKIAPFARITAWRFTVWLLHDEYDYVKDKKYLSVTTLLKPIKHIVMKHRVDLSVESMD 4834
Qy 352 ---------GTGL--VGEKYWQAIR--ALRKAG----------------------IPM 374
Db 4835 IMDLVSTSMGTGLHDSIEKAWHNGHHNALRKLGYPERVVQNVVINPTPEQIASNPNLIPV 4894
Qy 375 --------------------VVATG------------------------------------ 379
Db 4895 WIEQRATKKIAGWTIGGKFDIVTEGLLQDVKSTSTYTWVKGGRDEEHRLQGSMYRWLHDD 4954
Qy 380 ---------NY---------ATSA------------------SSSSW------------ 390
Db 4955 KITEDVIRINYIFTDWMKALATSQENYPKHRVLHKDIPLMSYEATEEWIKAKLALIDKYW 5014
Qy 391 -- 390
Db 5015 DAPESEIPECTDEELWRSDPVFKYFADPEKAKQPGARSTKNFPTLFEARKFMAEKGKGTI 5074
Qy 391 ----------DLVANNHLKM-------------TDTG---------------------- 404
Db 5075 LHVPEIFPMIDLTGVSHHPAIEEIVDVLCNKTQNTDRGFFRVEVAYFLAKMASCMGATII 5134
Qy 405 ---------NVTRTA--------AH------------------EDAI-AVA--------- 419
Db 5135 TKDRGEIPVNIYAVALATSGFGKGHSVNIVEDGFMAGFKKRFMEDTMPALAEDRLWKIAN 5194
Qy 420 --SAKN---QTVEFDKVNIGGESFKYRNIGAF---FD---------------------- 448
Db 5195 ERSARNATDQKEEFDKV----EA-EYRRTGAYPFTFDSGTPPAVKQLRHKLLLAGCGAIN 5249
Qy 449 ----------------------------KSKITTN--------E-DG-TKA-------P 462
Db 5250 LQIDEIGSNLLANADVLTVFLELYDQGKIKQKLTKNTAESIRSEEVDGKTPANLLLFGTP 5309
Qy 463 SKL-------KFVY--------------IGKG-------------------QD------- 475
Db 5310 SKLLNGSQTEELFYDFLDTGYSRRCLFAVGQADKRAYLSQSAEEIYRNLIKQDNNAVVNK 5369
Qy 476 --------------------QDLIG------------------------------------ 480
Db 5370 WANHFHSLADASKFGWKMTVEDDVGIALIDYKIQCEKQAAALADHEEIRKAELEHRYFKA 5429
Qy 481 LDLRGKIAVMD----------------------------RIYTK------------- 496

```
Db       5430 LKLAGAFAFVDNSIMVEMEHLKQAILLVEESGVAFQTILNREKAYVKLAKYIASVGTEVT 5489

Qy        497 --DL---------KNA----------------------------FK----KAMD---- 507
               ||          ||                            |:    |  |
Db       5490 HADLLEALPFYKSGNAARNEMMTLATAWGYKQHIIIKKTYNEGIEFFRGETLKETDTNEM 5549

Qy        508 ---------------------------------------KGARA--------IMVVNTVN 520
                                                     :| ||         |:|  |:
Db       5550 IVSYGENFAYDYIAEKVPFDQLHVLTQQAGYHWANHHFRRGHRAEENAIAGFNMIVIDVD 5609

Qy        521 --------------Y----Y---------NR----------------------DN---WT 528
                           |    |         ||                      ||   |
Db       5610 GTCSLDVCHELMKEYRFMTYTTKRHTDEENRFRLIIPMNYQLELDGAEYKEFMDNVMAW- 5668

Qy        529 ELPAMGYEADE--------------GT--------------------KSQVF-----SIS 549
               ||   :::||               ||                    | ::|      :
Db       5669 -LP---FKTDESANQRAKKWESCAKGTYYYNMDAPLLDVRDFIPRTSKNEAFQQGMQKVQ 5724

Qy        550 GDDGVKLW-----------N-MI---------------------------NP---DK-- 564
               | :: |           | ||                           ||   |:
Db       5725 NLDNLERWFAQRIATGNRNNHMIKYALALVDSGIGFNEVKQRVHEFNKKLSNPLSADELD 5784

Qy        565 ------------------------------------------------------------ 564

Db       5785 STVMITVAKLLFLVEEDCQLKEETMSEEFEHGEMNDQLILIVGFSAAGKSASLRNIRNQE 5844

Qy        565 -----KTEV-KR----NNK------------EDF--------------------KDKLE 581
                    ||: ||    ||             | |                    :|:||
Db       5845 RWVYLNTEAGKRLPFKNNFNGVRITDPYQILEYFDQCIEHRDDVDGIIVDSLTFMMDMLE 5904

Qy        582 -QYY-----------------------------PI--------------DMESF---- 591
               ||:                            |:              ::::|
Db       5905 TQYVLTAANTQKAWGEFAQFFKILLQEKVVKFAKPVIFTAHVKDEVDERAMELKTFVPVK 5964

Qy        592 -----N----------------------SN--------------------KP------- 596
                   |                      ||                    :|
Db       5965 GSLKNNGIEAYFSTVVSAERIDLKELEKYSNGMLEITEDEQELGYKHVFQTRPTKKTVGK 6024

Qy        597 ---------------------------------------------NVG------------ 599
                                                          |:|
Db       6025 RIRSPMGMFSKQETYMDNDAQKLLDHLAEYYSTGQAAKAEKATDNLGGGFGAKESDIYLA 6084

Qy        600 ------------------------------------------------------------ 599

Db       6085 NLKVAYAGKAASGANFIQIIADLTDLDGHSAGEYREQLYITSGTEKGCKCTYEKNGKEYF 6144

Qy        600 --------D---------------EKEI----DF-------------------KFA---- 609
                       |               |:::    ||                   |||
Db       6145 LPGYTVINDILVMTSGETIPEAVFEEKVVNVYDFDEKKEVAKSVMVPVNAIGGKFAVAIL 6204

Qy        610 --------------------------------PDT-------------DKELYKEDIIV- 623
                                             ||                |||  |: ::
Db       6205 KSEEDKQTKDGSGNYVSTGETRFTNTIEKVFHPDLHLTVVEAEELTERGKELTVEEAVFW 6264

Qy        624 ----------------------------P------AGSTS------------------- 629
                                           |      ||:::
Db       6265 DKWLEKNKGVTRDKTTKGGASGKAGQPPKPGATNTGAGASAAKSLFEIMQIPVLGVDPSF 6324

Qy        630 --WG-PR------------IDLLL---KPDVSAPG-----KNIKS-------TLNVI--- 656
                 ||  |            :|| |   |||    |      | |           |  ||
Db       6325 RNWGLARGMLDLETGILSGLDLKLVETKPD----GTKQVRQNSKDMQAAEDITTGVIDWF 6380

Qy        657 -----------------NGKSTYGY 664
                              ||: :||:
Db       6381 KEAKVIFVEVPVGSQSANGMKSYGV 6405
```

RESULT 10
ABG23295
ID ABG23295 standard; Protein; 5909 AA.
XX
AC ABG23295;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23286.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS87482.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 53654; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5909 AA;

Query Match 26.7%; Score 1173; DB 22; Length 5909;
Best Local Similarity 9.2%; Pred. No. 2.7e-07;
Matches 311; Conservative 169; Mismatches 153; Indels 2730; Gaps 130;

```
Qy          1 YPVVLADTS-----------------SSEDALN---------ISD-------KE----- 21
                | | ::|:                 :|| : |         : |       ||
Db       2305 YAVWITDVDPRNHRIEVIHVIREISGRTSEYVNNLLAQPTPFKVFDWLTETQAKEFVRRL 2364

Qy         22 ----------------------KVAE--NKEK--------------------H----- 30
                                    |::|  ||||                    |
Db       2365 NNTQGVISVYNKLEDGIPVKVAVKATEPTNKEKLEAASTKLKAMGAVATDAYEFHTDLNE 2424

Qy         31 ----------------ENIH-----SA-------------METSQD-----------F 43
                              :|:|     ||             : |  |           |
Db       2425 NEETIVAVVNDKVVPDVQNLHRHIRQSANLKDYKGFTKFLERLSTVIDKRRHSVEDLMKF 2484

Qy         44 KE---------------------------------------------------------- 45
```


Db 2485 MEKGDLPIADDGSIVIFKRLESKDNPEHLKDVMDDVFVDCHSKKIEQCVGMRVMVKENLV 2544
Qy 46 ---------- 45
Db 2545 DQNRRKDCSHGLHVASLQYIRNFSGNVTIIGKVAPEDVFAVPEYDVTKMRVSAYHIITKL 2604
Qy 46 ---------- 45
Db 2605 PDAIRDHVNNGNPISTIEGGTEILNMVLSGNHPEPSRQVLVGGHYGTNLTYSELTRNYLD 2664
Qy 46 ----------KKTAVIKEKEVVSKNP----------VIDN 65
Db 2665 DDEDSVIPDEPVIKKTALNMEESFTATNPKAEPVKATDVKPIKKTKMEILKDLWNQFLKA 2724
Qy 66 NTSNE---------EAKIKEENS----------NK---------- 81
Db 2725 ETSNEAITLADEIIAAKGKCKKSWASLGFSGDMVQKLIDARANKPKPEKAVKQDKTVAPK 2784
Qy 82 ----------SQGD---- 85
Db 2785 KTTVTRSRNADVIRSYLNDKGMSDYCKAHAIHDVKRAAKKSYAALGLTPEECSQIDKLKH 2844
Qy 86 ----------Y-------TDSFVNKNT---------- 95
Db 2845 HLKIHTMAKVYRANRKAHDDDIIRMNAVGLSLGTIAKTLGVHPTTVTLRLRSLGIEPADT 2904
Qy 96 -----EN----------PKKE---------- 101
Db 2905 RRTFMENVLAPLPTHVADWLSEQVGPKYEIRQYVRDLLLEAYNNRHNHKETAHERFLKRY 2964
Qy 102 ----------DKVVYI---------- 107
Db 2965 VEVVEMLEELAPNSEEAYDTLHNAIAALTELSELMKSSDKEVYAPNSNRLAILDALADQI 3024
Qy 108 ----------AEFKDKE----SGEKAIKELS---------- 124
Db 3025 VTATGIGTFLGMNVPGALTEVNRSNYSKFEDGEPVFNENKKVMKEAAAEGFFDFLLDPNA 3084
Qy 125 ----------SLKNTKVLY---------- 133
Db 3085 SELRISGPGGTGKTFTMAHMIDEIMPRYHETCSLMGIPALYNEVIMTATTNKAAEVLARA 3144
Qy 134 ------TY----------DR--------- 137
Db 3145 TGRPTSTYHSFQGLTVRNNLKTGEADLIPSKSFSIKKNKVIFIDEASMIDRKLKKFINEG 3204
Qy 138 ----------IFNGSA---------- 143
Db 3205 THQSKIVYLGDKCQLLPVKETSSPVYNDSSIKEFMLTQQMRTDVPELHALHEQLRGTIEG 3264
Qy 144 ------IETTPDNLDKI-------KQIEG---------ISSVERAQKVQ-----PMMN-- 174
Db 3265 KHGFLPIKAVPGIIDWIKDGDEMRKEVEGHFMTKTDTRIVAYTNQQVVNYNTFIRAMNGF 3324
Qy 175 HARKEIGVEE----------AIDYLK----------SIN 193
Db 3325 HGEFCIG-EELVSNSAIQIGAGDRLSIEQECQIVDQDSATRKILIDSSGIELEVRDCTLD 3383
Qy 194 APFGKNFDG----------RGMVISNI--------DT----GTDY---------- 216
Db 3384 TGYGGVFEGIPIPKEPFPELRALHASTVHKSQGSTYDTIFIDATDLSTCRQPDVVARLLY 3443
Qy 217 ------RHKA----------MR--------- 222
Db 3444 VAVSRARHRVVFYGSLAEKFGGEFYSSSRSSQQPPSGHRLALANHLLDDMREYLEQAGKL 3503
Qy 223 ----------IDDDAKA 229
Db 3504 LLEVHMVNQMVFRLVQGCKTYQDVRDALPECLIAMDKQGHLKSLQRTRKAAFTLDNDAMA 3563
Qy 230 ----------SMR-----FK----------KEDLKGTDKNYWLS 248
Db 3564 LRHMRKSCRLSNTMREPILSFEVCMRYITFNNEGIYPIAILTTTLRKDDIVKAYMEPWDV 3623
Qy 249 DK----------IPHAF----------NYY------- 258
Db 3624 DKDSVMILELHTAPGKKKTPAKEMQEFITQELVP-AFEAAQTQYIICGDSEYYKQLTKQG 3682
Qy 259 ----NGG----------KIT----------VEK--- 267
Db 3683 KADANIGYVMDCAYGNQKVIYVPNYRQIFYDPDKIKTKIALSMQALIDYVSGSYVEPGTK 3742
Qy 268 ----------YDDG---------- 271
Db 3743 IIHYADYPLTPEAISKWLDKLLEMDVPLAVDIETFSLKHYDCGIGTITFCWNKHEGIAFP 3802
Qy 272 ----------RDYF---------- 275
Db 3803 VDYEPIEGATEAPYGRQVHNMLVRSMLRDFFIKYLNRQMYHNIAFDVYALIYQLFMTDLL 3862
Qy 276 DP----HGMHI----------AGI----------LAGN---------- 289
Db 3863 DTEGLLHGMSIMLRNWDCTKLITYLATNSCAGNKLSLKDQAQEYAGNYAQEEINDITRIP 3922
Qy 290 ----------DTE---------- 292
Db 3923 LAELLEYNLVDGLCTWYVYEKHWDTLVNDQQLDVYTNIFKPACEDIIQMQLTGMPINMRT 3982
Qy 293 ---------QDIKN----------FNG 300
Db 3983 VLEVEEALTQDYNNALKTIADSKVIKDFTRLLNEEWVEKQNQILKKKRVTLADAKEQFNP 4042
Qy 301 IDGIA---------- 305
Db 4043 NSGIQLQKLLFEFLGLPVLGLTASKLPATGSGILKSLKNHTQDTSILEILDALIDYKAVD 4102
Qy 306 ----------PNAQ---------- 309
Db 4103 KILTAFIPALKNARQGPDGWHYLFGNLNLGGTVSGRLSSSEPNLQNLPSGSRYAKMIKKC 4162
Qy 310 -------IP----------SYKMYSDAGSGFAGDETMFHA--------- 332
Db 4163 FEAPPGWIFCGLDFASLEDRISALTTKDPNKLKVYTD---GYDG-----HSLRAYAYFGE 4214
Qy 333 ----IEDSIKHNVDVVSVSS----------GF--TGTG 354
Db 4215 MMPDIEDTVE------SVNSIQEKYKAYRQDSKAPTFALTYQGTYITLMKNCGFPEQKAR 4268
Qy 355 LVGEKY---------W----------QAIRALRKAGIPMVV- 376
Db 4269 MVEERYHTLYKVSDDWVQAKLDQAAKDGYVTVAFGLRVRTPLLQQVIRGTSKT--PYEAE 4326
Qy 377 ----ATGNYATSASSSSW---------- 390
Db 4327 AEGRTAGN----ALGQSWCLLNNRAGSEFMRKVRNSKHRLDIRPSTHIHDAQYFLIRDDM 4382
Qy 391 DLV--ANNHLKMT----DTGNVTRTAAHED----AIAVASAK----------NQT 425
Db 4383 DVVIYTNTHLVKAVQWQDHPDIAHPDVHLDGQEGAVIVQNTMLLTEEKQVTFRDLGRAQS 4442
Qy 426 VEFDKVN----------IG---------- 434
Db 4443 ALFQNLNDRFGQVVDVRDIVFLNVSYLGHMSQPDFQQNLVTNPQRMEQPPYQAECEDIIQ 4502
Qy 435 --------GESFK--YRN----IG-------AFFDKSK-------ITTNEDG-----TKA 461
Db 4503 ALNMTFDEGCLFKALWRNAAARLGNGKPGNNAVYDAEKMVHYANRIFIKESGGQKKEAEE 4562
Qy 462 P----------SKLKFVYIG---KGQDQDLIG---------- 480
Db 4563 PVSAALKKGEEIHMSLSELFTLLTGEAAEGEVMSLAGWHTHVPNRLGVMPIAVAMSKAEQ 4622
Qy 481 ----------LDLRGKIA----------VMDRIY------- 494
Db 4623 VDVVRADGTTITHVFPHEIDWSDKIAPFARITAWRFTVWLLHDEYDYVKDKKYLSVTTLL 4682

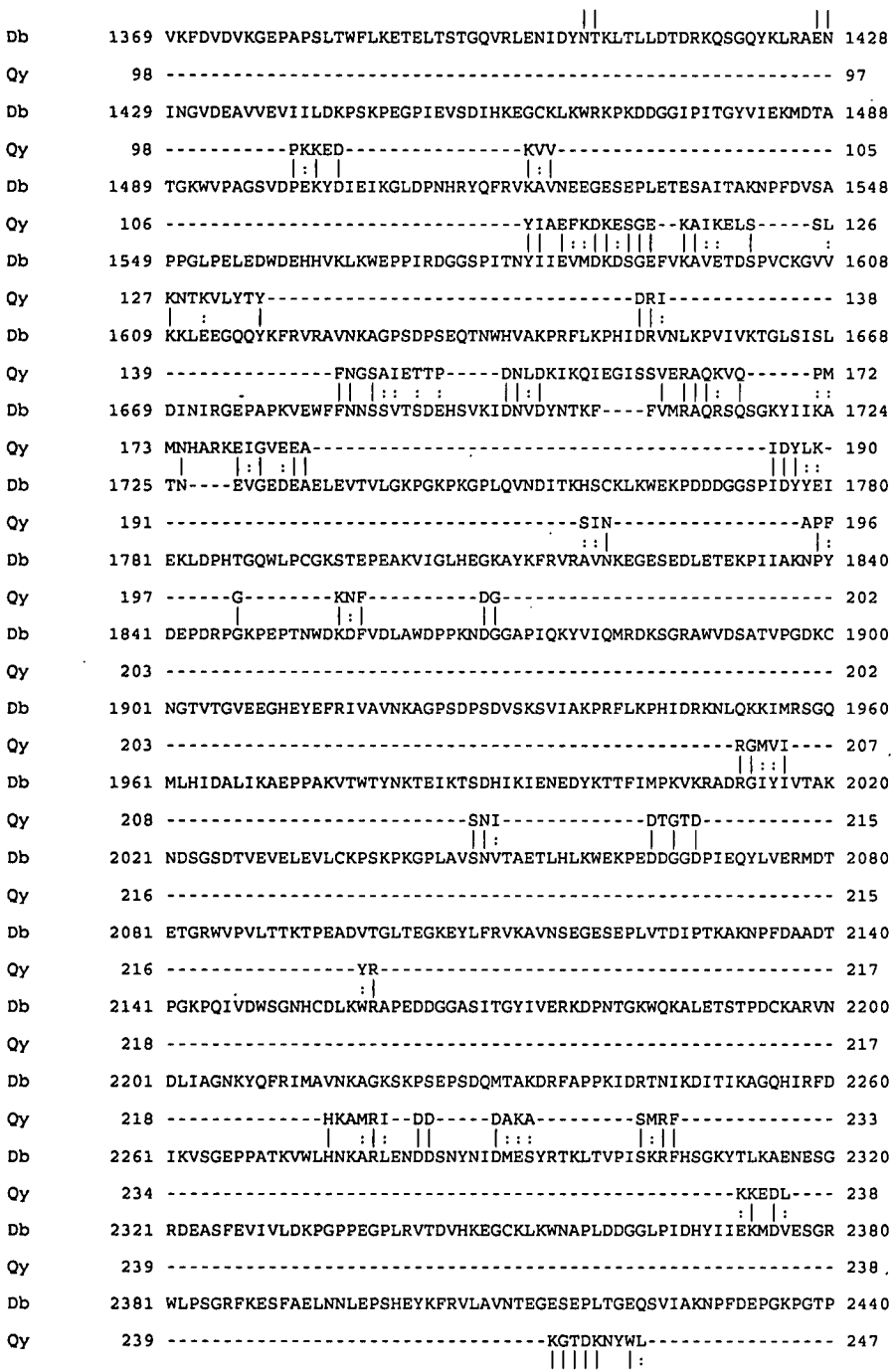
```
Qy        495 ----------:----------TKDLKNAFKKAMDKG----------------------ARAI- 513
                         ::|::: :  :|  |                              |::
Db       4683 KPIKHIVMKHRVDLSVESMDIMDLVSTSMGTGLHDSIEKAWHNGHHNALRKLGYPERVVQ 4742

Qy        514 -MVVNT---------------------------------------------------- 518
               |:|
Db       4743 NVVINPTPEQIASNPNLIPVWIEQRATKKIAGWTIGGKFDIVTEGLLQDVKSTSTYTWVK 4802

Qy        519 ----------------------------VNYYNRDNWTE------------------LPA 532
                                          :||   | | :                  :|
Db       4803 GGRDEEHRLQGSMYRWLHDDKITEDVIRINYIFTD-WMKALATSQENYPKHRVLHKDIPL 4861

Qy        533 MGYEA--------------------------DE------------------------- 539
              | |||                         ||
Db       4862 MSYEATEEWIKAKLALIDKYWDAPESEIPECTDEELWRSDPVFKYFADPEKAKQPGARST 4921

Qy        540 -----------------GT------------------------------KSQ-----V 545
                              ||                              |:|     :
Db       4922 KNFPTLFEARKFMAEKGKGTILHVPEIFPMIDLTGVSHHPAIEEIVDVLCNKTQNTDRGF 4981

Qy        546 FSI-----------------------------------SG----------DDGVK---- 555
               | :                               ||          :||::
Db       4982 FRVEVAYFLAKMASCMGATIITKDRGEIPVNIYAVALATSGFGKGHSVNIVEDGFMAGFK 5041

Qy        556 --------------LWNMINPDKKTEVKRNN---KEDFKDKLE-QY-----YPIDMESF 591
                           || : |:       ||    ||:| ||:| :|     ||    :|
Db       5042 KRFMEDTMPALAEDRLWKIANERS----ARNATDQKEEF-DKVEAEYRRTGAYPF---TF 5093

Qy        592 NSNKP-----------------N-----VG-----------------DE---------- 601
               :|  |                |     :|                 |:
Db       5094 DSGTPPAVKQLRHKLLLAGCGAINLQIDEIGSNLLANADVLTVFLELYDQGKIKQKLTKN 5153

Qy        602 -------KEID-------FKF-AP-----------------DT-------------DK-- 614
                     :|:|       | | :|                 ||             ||
Db       5154 TAESIRSEEVDGKTPANLFLFGTPSKLLNGSQTEELFYDFLDTGYSRRCLFAVGQADKRA 5213

Qy        615 -------ELYK---------------------------------ED------------- 620
                     |:|:                                 ||
Db       5214 YLSQSAEEIYRNLIKQDNNAVVNKWANHFHSLADASKFGWKMTVEDDVGIALIDYKIQCE 5273

Qy        621 -----------------------------------IIV------------------- 623
                                                  |:|
Db       5274 KQAAALADHEEIRKAELEHRYFKALKLAGAFAFVDNSIMVEMEHLKQAILLVEESGVAFQ 5333

Qy        624 --------------------------------P---AGSTS----------------W 630
                                             |   :| ::                |
Db       5334 TILNREKAYVKLAKYIASVGTEVTHADLLEALPFYKSGNAAQKVPFDQLHVLTQQAGYHW 5393

Qy        631 -------GPR-------------ID----LLLKPD------------------------ 641
                    |:|             ||    :|:::|
Db       5394 ANHHFRRGHRAEENAIAGFNMIVIDVDVLFLVEEDCQLKEETMSEEFEHGEMNDQLILIV 5453

Qy        642 -VSAPGK-----NI-------------------KSTLN---------------------- 654
                :|| ||     ||                    | ::|
Db       5454 GFSAAGKSASLRNIRNQERWVYLNTEAGKRLPFKNNFNGVRITDPYQILEYFDQCIEHRD 5513

Qy        655 ----------------------------------------------------VI----- 656
                                                                   ||
Db       5514 DVDGIIVDSLTFMMDMLETQYVLTAANTQKAWGEFAQFFKILLQEKVVKFAKPVIFTAHV 5573

Qy        657 ----------------------NGKSTY------------------------------- 662
                                    ||  :|
Db       5574 KDEVDERAMELKTFVPVKGSLKNNGIEAYFSTVVSAERIDLKELEKYSNGMLEITEDEQE 5633

Qy        663 -GY 664
               ||
Db       5634 LGY 5636
```

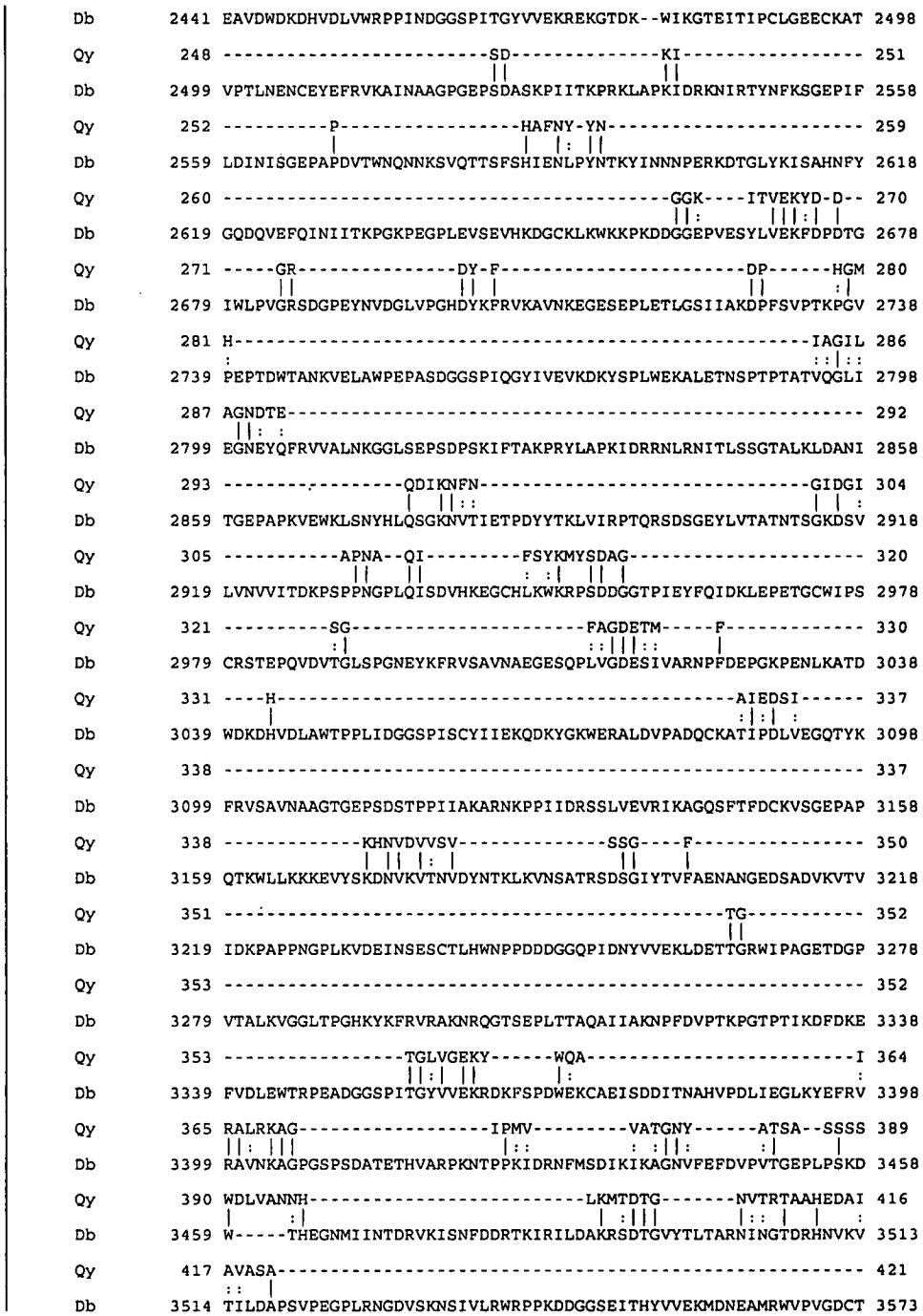
RESULT 11
ABB58144
ID ABB58144 standard; Protein; 7107 AA.

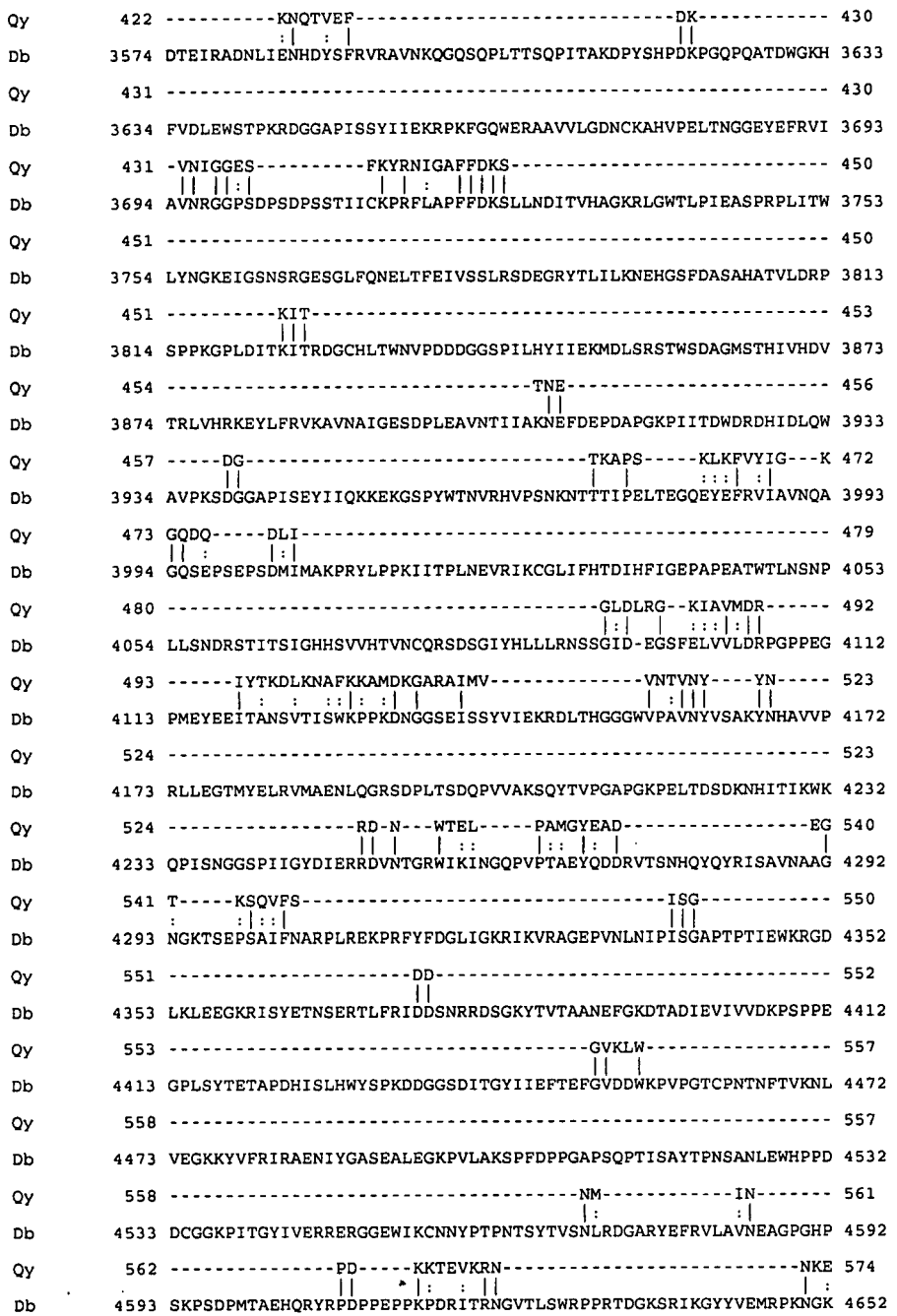
XX
AC ABB58144;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1224.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02247.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7107 AA;

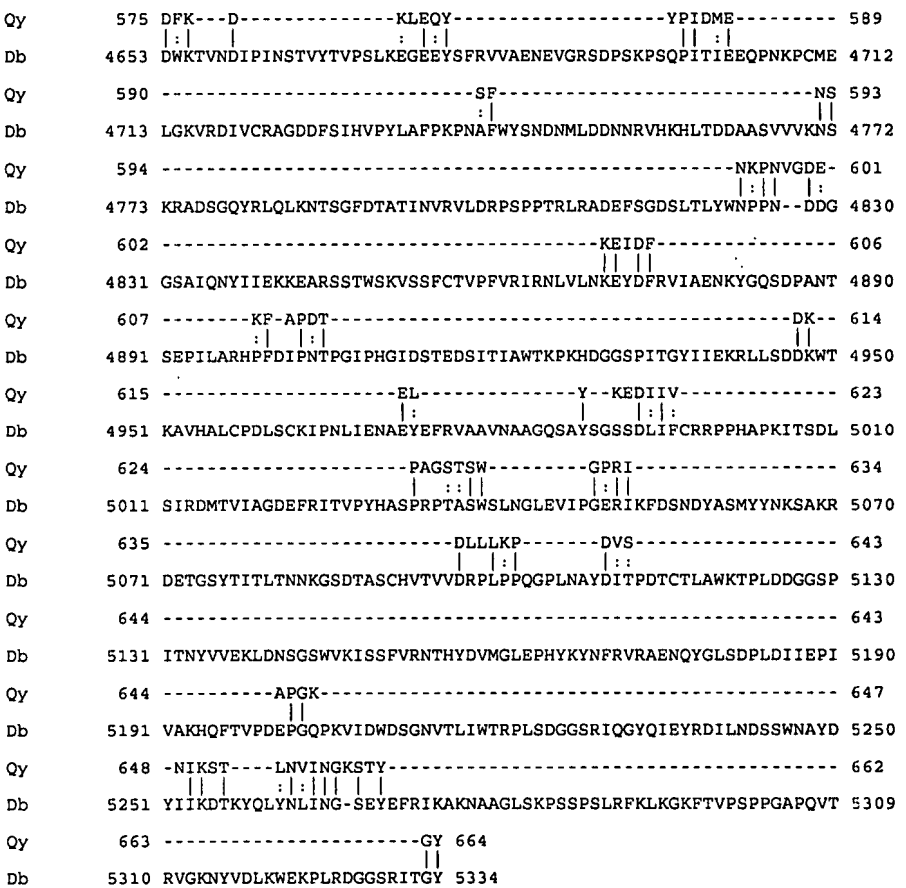
Query Match 26.6%; Score 1167.8; DB 22; Length 7107;
Best Local Similarity 7.6%; Pred. No. 7e-07;
Matches 324; Conservative 163; Mismatches 156; Indels 3642; Gaps 139;

```
Qy          1 Y---------PVVL---AD--------TSSSEDALNISDKEKVA--ENKEKHENI----- 33
              |         ||:     |        ||||   ||    |::  |:  ||  |
Db       1071 YDVNVRGEPAPVITWYQNDKELKPEELPSSSE-IKNIPYNTKISIIETVRKHTGIYKIIA 1129

Qy         34 ---HSAMETS---------------------------------QD-------------FK 44
                 | : |::                                :|             |
Db       1130 VNEHGQDEATVEVNILAPPSKPRGPLDVKDVTKDSCKLKWKKPEDDGGKPISAYQVEKFD 1189

Qy         45 EKK-------------------------------TAVIKEKE---------VVSKNP--- 61
              :|                               |: :| |          :::|||
Db       1190 KKQGRWVPLGRTSANDTEFDVKGLQEGHEYQFRVKAINEEGESDPLDSDDSIIAKNPYDA 1249

Qy         62 --------VID-------------------------------------------NNTSNEEA 72
                    ::|                                          :||| :||
Db       1250 ASKPGTPNIVDYNEHMVKLKWEAPRSDGGAPISGYIIEKKDKFSPIWDEILSTNTSVPEA 1309

Qy         73 KIK---EEN--------SNKSQG--DYTDS-------------FVNK------------- 93
                ::   | |          || :|  | :|:             ::|:
Db       1310 TVEGLVEGNIYQFRVRAVNK-AGFSDPSDATEPHLAKPRNLKPYINRDKMKPIKVRAGQP 1368

Qy         94 -----------------------------------NT------------------EN 97
```









```
RESULT 12
ABB98574
ID   ABB98574 standard; protein; 2835 AA.
XX
AC   ABB98574;
XX
DT   14-JAN-2003  (first entry)
XX
DE   Dextran saccharase, DSRE.
XX
KW   Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
KW   DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
KW   microflora regulation; intestinal transit; mineral assimilation;
KW   colon cancer; acne; dandruff; body odour.
XX
OS   Leuconostoc mesenteroides NRRL B-1299.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..40
FT                   /label= Signal_peptide
FT   Protein         41..2835
FT                   /label= Mature_protein
FT   Peptide         423..439
FT                   /label= SEQ_ID_6
```

```
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         478..501
FT                   /label= SEQ_ID_7
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         519..539
FT                   /label= SEQ_ID_8
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         560..571
FT                   /label= SEQ_ID_9
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         631..645
FT                   /label= SEQ_ID_10
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         1014..1021
FT                   /label= SEQ_ID_11
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2120..2138
FT                   /label= SEQ_ID_12
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2161..2184
FT                   /label= SEQ_ID_13
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2202..2214
FT                   /label= SEQ_ID_14
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2243..2250
FT                   /label= SEQ_ID_15
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2315..2322
FT                   /label= SEQ_ID_16
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Peptide         2689..2696
FT                   /label= SEQ_ID_17
FT                   /note= "This sequence is specifically claimed in Claim 7"
FT   Domain          1981..1142
FT                   /note= "Catalytic domain, SEQ ID 1. This sequence is
FT                   specifically claimed in Claim 4"
XX
PN   FR2822163-A1.
XX
PD   20-SEP-2002.
XX
PF   19-DEC-2001; 2001FR-0016495.
XX
PR   16-MAR-2001; 2001FR-0003631.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Bozonnet SAM,  Remaud SMMC,  Willemot RML,  Monsan PEF;
XX
DR   WPI; 2002-715213/78.
DR   N-PSDB; ABQ80961, ABQ80962.
XX
PT   New glycosyl transferase enzymes, containing glucan bonding and
PT   catalytic domains and producing alpha-(1-2) branched dextrans, useful
PT   in probiotic, pharmaceutical or cosmetic compositions  -
XX
PS   Claim 6; Page 65-74; 82pp; French.
XX
CC   The present sequence is a novel dextran saccharase, DSR-E, from
CC   Leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl
CC   transferase activity suitable for producing dextrans having alpha(1-2)
CC   branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,
CC   alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbopyranoside or
CC   alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in
CC   prebiotic, pharmaceutical or cosmetic compositions. The dextrans and
CC   related compounds having alpha(1-2) bonds, produced using DSR-E, may be
CC   involved in signalling/cellular recognition processes in vivo
CC   (specifically in regulation of microflora in the intestines or on the
CC   skin); and are potentially useful for improving intestinal transit,
CC   increasing assimilation of minerals (e.g. calcium and/or magnesium),
CC   preventing cancer of the colon and combating skin problems such as acne,
CC   dandruff and body odour.
```

```
XX
SQ   Sequence   2835 AA;

  Query Match             26.5%;  Score 1165.4;  DB 23;  Length 2835;
  Best Local Similarity   10.8%;  Pred. No. 2.2e-08;
  Matches  285;  Conservative  160;  Mismatches  202;  Indels 1986;  Gaps  120;

Qy          2 PVV--LADTSSSEDA---LNI--------------SDKEKVAENKEKH----------- 30
              ||:   |||||  |:    |:               :| : |  ::  |
Db        196 PVANKAADTSSIHDQPLDTNVPTDKSANLVSTTQKSTDNQQVKSTETSHLQEINGKTYFL 255

Qy         31 ------------------------------------------ENIHSAME---------- 38
                                                        :| | |:
Db        256 DDNGQVKKNFTAIIDGKVLYFDKTSGELTANAPQVTKGLVNIDNAHNAAHDLTADNFTNV 315

Qy         39 ------------------------TSQDF--------KEKKTAV--------------- 50
                                       |::||        ::| |:|
Db        316 DGYLTANSWYRPKDILKNGTTWTPTTAEDFRPLLMSWWPDKNTQVAYLQYMQSVGMLPDD 375

Qy         51 ----------------------IKEKEVVS---------KNPVID-------------- 64
                                    |: :  ||         :|::||
Db        376 VKVSNDDNMSTLTDAAMTVQKNIESRIGVSGKTDWLKQDMNKLIDSQANWNIDSESKGND 435

Qy         65 ------------------------NNT-SNEEAKIKE--------------ENSNK-- 81
                                      | | :|:   | :              :|||:
Db        436 HLQGGALLYVNDDKTPNANSDYRLLNRTPTNQTGQITDPSKQGGYEMLLANDVDNSNPVV 495

Qy         82 -----------------SQGDYTDSF-------------------------------V 91
                               :| | |  |                               :
Db        496 QAEQLNWLHYMMNIGTIAQNDPTANFDGYRVDAVDNVDADLLQIAGDYFKAAYGTGKTEA 555

Qy         92 NKN------------------------------------------------------- 94
              | |
Db        556 NANNHISILEDWDNNDSAYIKAHGNNQLTMDFPAHLALKYALNMPLAAQSGLEPLINTSL 615

Qy         95 -------TEN---PKK-----------------EDKV------------------VYI 107
                  |||   |                   :||:                    :|
Db        616 VKRGKDATENEAQPNYAFIRAHDSEVQTVIAQIIKDKINTKSDGLTVTPDEIKQAFTIYN 675

Qy        108 A-EFK-DKESGEKAIKE----LSSLKNT--KVLY----------------TYD------- 136
              | |:| |||     |:     | : |:|  :|:|                 ||
Db        676 ADELKADKEYTAYNIPASYAVLLTNKDTVPRVYYGDLFSDDGQYMSQKSPYYDAITSLLK 735

Qy        137 -RI-----------------FNGSAIETTP----------------DNL----------- 151
               ||                 |: :  || |               |:|
Db        736 SRIKYVAGGQSMNMTYLHECFDPAKNETKPQGVLTSVRYGKGAMTADDLGNSDTRQQGIG 795

Qy        152 ------------------------------------------------------------ 151

Db        796 LVINNKPFLNLNDDEQIVLNMGAAHKNQAYRPLMLTTKSGLQIYDKDAGAPVVYTNDAGQ 855

Qy        152 -----DKI-----KQIEGI-----------SSVERAQKVQ---------------------- 170
                  |::     :|: |            |   |:|  |
Db        856 LIFKSDMVYGVSNPQVSGYFAAWVPVGASDSQDARTQSSQSETKDGDVYHSNAALDSNVI 915

Qy        171 ----------PMMN------------------------------------------------ 174
                       | :|
Db        916 YEGFSNFQAMPEKNDDFTNVKIAQNAKLFKDLGITSFELAPQYRSSTDNSFLDSVIQNGY 975

Qy        175 ----------------------------HARKEIGVEEAIDYLK---------------- 190
                                          . ||:   |::   |:::
Db        976 AFTDRYDVGYNTPTKYGTVDQLLDSLRALHAQ---GIQAINDWVPDQIYNLPGEQIVTAV 1032

Qy        191 ------------------------------------------------------SINAP- 195
                                                                       | |:|
Db       1033 RTNGSGKYDYDSVINNTLYDSRTVGGGEYQEKFGGLFLDQLKKDYPSLFETKQISTNQPM 1092

Qy        196 -------------F-GKNFDGRGM--VIS--------NI--DTG---------------- 213
                          | | |  |||:  |:         |:  |:|
Db       1093 NPDVKIKEWSAKYFNGSNIQGRGAWYVLKDWATNQYFNVSSDNGFLPKQLLGEKTSTGFI 1152

Qy        214 -----------TDYRHKAMRIDD-------DAKASMRFKKEDLKGTDKNYWLSDKIP--- 252
                        : |: |   | |       |    |   |:::   ||||::   :|
Db       1153 NPDVKIKEWSAKYFNGSNIQGRGAWYVLKDWATNQYFNVSSDNGFLPKQLLGEKTSTGFI 1212
```

```
Db       1153 TENGKTSPYSTSGYQAKDTFIQDGTNWYYFDNAGYMLTGKQNI--HDKNYYF---LPNGV 1207

Qy        253 -----HAFN-----YYN--GGKITVEKYDD-----------GR-----------D----- 273

Db       1208 ELQDAYLFDGNQEFYYNKAGEQVMNQYYQDSQNQWHYFFENGRMAIGLTEVPNADGTHVT 1267

Qy        274 -YFDPHGMHIAG-------------------------------------------ILAGN-- 289

Db       1268 QYFDANGVQIKGTAIKDQNNQLRYFDEATGNMVVNSWGQLADKSWLYLNAQGVAVTGNQK 1327

Qy        290 -DTEQ-----DIK---------------------------------------------- 296

Db       1328 IDGEEYYFNADGKQVKGNAIIDNNGDQRYYDGDKGVMVVNSWGELPDGSWLYLNDKGIAV 1387

Qy        297 ----------NFNGIDGIAPNAQIF-SYKMYSDAGSGFAGDE-------------TMF- 330

Db       1388 TGRQVINNQVNFFGNDG----KQIKDAFKLLSD-GSWVYLDDKGLITTGAKVINGLNMFF 1442

Qy        331 ----HAIE-----DSI--KHNVD-----VVSVSSG---------------FTG------ 352

Db       1443 DKDGHQIKGDASTDANGKRHYYDKNDGHLVTNSWGELPDGSWLYLEEQGDAVTGQRVIDG 1502

Qy        353 ---------------------------------TGL--VGEK--Y-------------- 360

Db       1503 KTRYFDEDGKQIKNSLKTLANGDKIYLDGDGVAATGLQHVGDKIMYFDEDGKQVVGKFVS 1562

Qy        361 -----W-----------------------QAIRALRKAGIPMVVATGNYATSAS----- 386

Db       1563 AKDGSWYYLNQDGVAAVGPSSINGQSLYFDQDGKQVKYNEVRNSDGTTNYYTGLTGEKLT 1622

Qy        387 -------SSSW---DL---------VAN---------------------NHLKM--TDT 403

Db       1623 QDFGELPDGSWIYLDAQGHTVTGAQIINGQNLYFKADGQQVKGHAYTDQLGHMRFYDPDS 1682

Qy        404 GN--------------------VTRTAAHE--------DAIAVASAKNQ-TVE-----F 428

Db       1683 GDMLSNRFEQITPGVWAYFGADGVAITGQHDINGQKLFFDETGYQVKGSQRTIDGTLYSF 1742

Qy        429 DK---------------------------------------------------------- 430

Db       1743 DSQTGNQKRVQTTLLPQAGHYITKNGNDWQYDTNGELAKGLRQDSNGKLRYFDLTTGIQA 1802

Qy        431 ----VNIGGES--FK---------------------------------YR---------- 441

Db       1803 KGQFVTIGQETYYFSKDHGDAQLLPMVTEGHYGTITLKQGQDTKTAWVYRDQNNTILKGL 1862

Qy        442 -NI-G--AFFDKSKITTNE--DGTKAPSKLKFVYI--GKG----------QDQDLIGLDL 483

Db       1863 QNINGTLQFFDP---YTGEQLKGGVAKYDDKLFYFESGKGNLVSTVAGDYQDGHYISQDG 1919

Qy        484 R-----------------GKIAVMD---------------RIYTKDLK---------- 499

Db       1920 QTRYADKQNQLVKGLVTVNGALQYFDNATGNQIKNQQVIVDGKTYYFDDKGNGEYLFTNT 1979

Qy        500 -----NAF---------------------------------------KKAMDKGARAIM 514

Db       1980 LDMSTNAFSTKNVAFNHDSSSFDHTVDGFLTADTWYRPKSILANGTTWRDSTDKDMRPLI 2039

Qy        515 VV-----NT-VNYYN--RDNWTELPAMGY--EADE-----------GTKSQVFSISGDD 552

Db       2040 TVWWPNKNVQVNYLNFMKANGLLTTAAQYTLHSDQYDLNQAAQDVQVAIERRIASEHGTD 2099

Qy        553 G---------------VK---LWNM--------------------INP------------ 562

Db       2100 WLQKLLFESQNNNPSFVKQQFIWNKDSEYHGGGDAWFQGGYLKYGNNPLTPTTNSDYRQP 2159

Qy        563 ------------DKKTEV----------------------------KRNNKEDF--KDK 579

Db       2160 GNAFDFLLANDVDNSNPVVQAENLNWLHYLMNFGTITAGQDDANFDSIRIDAVDFIHNDT 2219

Qy        580 LEQYY------------------------------------------------------ 584

Db       2220 IQRTYDYLRDAYQVQQSEAKANQHISLVEAGLDAGTSTIHNDALIESNLREAATLSLTNE 2279
```

```
Qy        585 -----P-------ID---------------------------------------------- 587

Db       2280 PGKNKPLTNMLQDVDGGTLITDHTQNSTENQATPNYSIIHAHDKGVQEKVGAAITDATGA 2339

Qy        588 -------------MESF-----------NS-NKPNV----------------GDE----- 601

Db       2340 DWTNFTDEQLKAGLELFYKDQRATNKKYNSYNIPSIYALMLTNKDTVPRMYYGDMYQDDG 2399

Qy        602 ----------------------------KEID-------FKF----------------- 608

Db       2400 QYMANKSIYYDALVSLMTARKSYVSGGQTMSVDNHGLLKSVRFGKDAMTANDLGTSATRT 2459

Qy        609 ---------AP-----DTDK-----------ELYKEDII---------------------- 622

Db       2460 EGLGVIIGNDPKLQLNDSDKVTLDMGAAHKNQKYRAVILTTRDGLATFNSDQAPTAWTND 2519

Qy        623 ---------------------------------VPAGST--------------------- 628

Db       2520 QGTLTFSNQEINGQDNTQIRGVANPQVSGYLAVWVPVGASDNQDARTAATTTENHDGKVL 2579

Qy        629 ----------------------------------------SWG-------P--------- 632

Db       2580 HSNAALDSNLIYEGFSNFQPKATTHDELTNVVIAKNADVFNNWGITSFEMAPQYRSSGDH 2639

Qy        633 --------------RIDLLLK-P-------------------------DV------SAP 645

Db       2640 TFLDSTIDNGYAFTDRYDLGFNTPTKYGTDGDLRATIQALHHANMQVMADVVDNQVYNLP 2699

Qy        646 GKNIKSTL---------------------------------------------------- 653

Db       2700 GKEVVSATRAGVYGNDDATGFGTQLYVTNSVGGGQYQEKYAGQYLEALKAKYPDLFEGKA 2759

Qy        654 ------NVIN-GKSTY---------------------------------GY 664

Db       2760 YDYWYKNYANDGSNPYYTLSHGDRESIPADVAIKQWSAKYMNGTNVLGNGMGY 2812
```

RESULT 13
ABG74786
ID ABG74786 standard; Protein; 31267 AA.
XX
AC ABG74786;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human RGS11 protein.
XX
KW RGS11; human; screening; cardiant; antianginal; gene therapy;
KW heart disorder; cardiac ischaemia; heart failure; angina.
XX
OS Homo sapiens.
XX
PN WO2002103355-A1.
XX
PD 27-DEC-2002.
XX
PF 17-JUN-2002; 2002WO-JP06019.
XX
PR 18-JUN-2001; 2001JP-0183038.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Koyama N, Tanida S, Yamamoto K;
XX
DR WPI; 2003-167557/16.
DR N-PSDB; ABX13540.
XX
PT Screening compounds regulating RGS11 expression and activity for
PT prevention and treatment of heart disease -
XX
PS Claim 1; Page 59-261; 321pp; Japanese.
XX

CC This invention describes a novel method for screening compounds for their
CC ability to regulate the activity and expression of human RGS11 and its
CC partial peptides and salts, by observing the expression or activity of
CC RGS11 in the presence or absence of the test compound. The products of
CC the invention have cardiant and antianginal activity and can be used for
CC gene therapy. The methods and compositions are useful in the prevention,
CC treatment and diagnosis of heart disorders such as cardiac ischaemia,
CC heart failure and angina. This sequence represents the human RGS11
CC protein described in the disclosure of the invention.
XX
SQ Sequence 31267 AA;

Query Match 26.5%; Score 1165.2; DB 24; Length 31267;
Best Local Similarity 8.8%; Pred. No. 0.00025;
Matches 312; Conservative 167; Mismatches 170; Indels 2908; Gaps 139;

```
Qy          1 YPVV--------------------LADT----SSSED---ALNI--SDKEKV------- 23
              :|||                    | |:    :|  |   :|||  :|:  :
Db       9448 LPVVFDCAISGSEPISVSWYKDGKPLKDSPNVQTSFLDNTATLNIFKTDRSLAGQYSCTA 9507

Qy         24 -----------------AENK--------------------EKH---------------- 30
                           :|:                      | |
Db       9508 TNPIGSASSSARLILTEGKNPPFFDIRLAPVDAVVGESADFECHVTGTQPIKVSWAKDSR 9567

Qy         31 -------------EN-IH-----------------SAMETSQD----------------F 43
                         ||  |                  :: |:  |                |
Db       9568 EIRSGGKYQISYLENSAHLTVLKVDKGDSGQYTCYAVNEVGKDSCTAQLNIKERLIPPSF 9627

Qy         44 KEKKTAVIKEKE---------VVSKNPVI----DNN-----TSNEEAKIK---------- 75
                :: : :::| |          |:   |:      ||      |||:|   |
Db       9628 TKRLSETVEETEGNSFKLEGRVAGSQPITVAWYKNNIEIQPTSNCEITFKNNTLVLQVRK 9687

Qy         76 ----------EENSNKS-----------------------------------Q------- 83
                      :: || :                                     |
Db       9688 AGMNDAGLYTCKVSNDAGSALCTSSIVIKEPKKPPVFDQHLTPVTVSEGEYVQLSCHVQG 9747

Qy         84 ----------------------------------------GDY---------TDS--- 89
                                                      |||         :|:
Db       9748 SEPIRIQWLKAGREIKPSDRCSFSFASGTAVLELRDVAKADSGDYVCKASNVAGSDTTKS 9807

Qy         90 ------------------------------------------------------FVN-- 92
                                                                    | |
Db       9808 KVTIKAEPIQFTKRIQNIVVSEHQSATFECEVSFDDAIVTWYKGPTELTESQKYNFRNDG 9867

Qy         93 ------------------------------------------------------------ 92

Db       9868 RCHYMTIHNVTPDDEGVYSVIARLEPRGEARSTAELYLTTKEIKLELKPPDIPDSRVPIP 9927

Qy         93 ------------------------------------------------------------ 92

Db       9928 TMPIRAVPPEEIPPVVAPPIPLLLPTPEEKKPPPKRIEESHERKVPAKVPEKKAPPPPKV 9987

Qy         93 ----------------------------------------------------KNT------- 95
                                                                  | |
Db       9988 IKKPVIEKIEKTSRRMEEEKVQVTKVPEIKPAIPLPAPEPKPKPEAGVPKKTPSPIEAER 10047

Qy         96 ---------ENPKKEDKVVY---------IAEFKD---KES---GEKAIKE-LSSLKNTK 130
                       | |: | :::|         : | ||    ||   |  || | | |  ::
Db      10048 RKLRPGSGGEKPPDEAPFTYQLKAVPLKFVKEIKDIILTESEFVGSSAIFECLVSPSTAI 10107

Qy        131 VLYTYDRIFNGSAIETTP------DNLDK------------------------------ 153
               : :  |     || |  :|      |  |:
Db      10108 TTWMKD----GSNIRESPKHRFIADGKDRKLHIIDVQLSDAGEYTCVLRLGNKEKTSTAK 10163

Qy        154 ----------IKQIE---------------------------------------GI-- 160
                       :| :|                                       |:
Db      10164 LVVEELPVRFVKTLEEEVTVVKGQPLYLSCELNKERDVVWRKDGKIVVEKPGRIVPGVIG 10223

Qy        161 -------------------SSVE------------------------RAQKVQPM---- 172
                              :||                          | | | |:
Db      10224 LMRALTINDADDTDAGTYTVTVENANNLECSSCVKVVEVIRDWLVKPIRDQHVKPKGTAI 10283

Qy        173 ---------------------------------MNH---------------------- 175

                                                ||
Db      10284 FACDIAKDTPNIKWFKGYDEIPAEPNDKTEILRDGNHLYLKIKNAMPEDIAEYAVEIEGK 10343

Qy        176 ---------ARK---------------------------------------------- 178
                        |:
Db      10344 RYPAKLTLGEREVELLKPIEDVTIYEKESASFDAEISEADIPGQWKLKGELLRPSPTCEI 10403

Qy        179 --EIG-------------------------------VEE-AIDY---LKSINAP------ 195
                 | |                             |:|  :|:   || :::|
Db      10404 KAEGGKRFLTLHKVKLDQAGEVLYQALNAITTAILTVKEIELDFAVPLKDVTVPERRQAR 10463

Qy        196 F--------------GKN-------FD----GRGMVISNIDTGTD--------------- 215
               |             |::       ||     |: :::    |:  |
Db      10464 FECVLTREANVIWSKGPDIIKSSDKFDIIADGKKHILVINDSQFDDEGVYTAEVEGKKTS 10523

Qy        216 -------------------------------YRHKAMRI-------------------- 223
                                            : |: | :
Db      10524 ARLFVTGIRLKFMSPLEDQTVKEGETATFVCELSHEKMHVVWFKNDAKLHTSRTVLISSE 10583

Qy        224 -------------DD---------------------------------------------- 225
                           ||
Db      10584 GKTHKLEMKEVTLDDISQIKAQVKELSSTAQLKVLEADPYFTVKLHDKTAVEKDEITLKC 10643

Qy        226 ----DAKASMRF-------------------------KKEDLK---------GTDKN--- 244
                |::: : |                          || |||         ||||:
Db      10644 EVSKDVPV-KWFKDGEEIVPSPKYSIKADGLRRILKIKKADLKDKGEYVCDCGTDKTKAN 10702

Qy        245 --------------Y-------------------------W---------------LSDKI 251
                           |                         |               : |
Db      10703 VTVEARLIKVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGK 10762

Qy        252 PHAFNYYN---G--GKITVE---------------------KYDD--------------- 270
                :| : : |    |  |:::::                 :: |
Db      10763 KHILILHNCQLGMTGEVSFQAANAKSAANLKVKELPLIPITPLSDVKVFEKDEAKFECEV 10822

Qy        271 -----------------GRDYF----D--PHGM--------------------HIAG--I 285
                              | | |    |  :| |                    | :|  |
Db      10823 SREPKTFRWLKGTQEITGDDRFELIKDGTKHSMVIKSAAFEDEAKYMFEAEDKHTSGKLI 10882

Qy        286 LAG-------------------------------------ND---------TEQD----- 294
                :|                                    ||         : ||
Db      10883 IEGIRLKFLTPLKDVTAKEKESAVFTVELSHDNIRVKWFKNDQRLHTTRSVSMQDEGKTH 10942

Qy        295 ----------------------------------------IKNFNGI------------ 301
                                                       : :::|:
Db      10943 SITFKDLSIDDTSQIRVEAMGMSSEAKLTVLEGDPYFTGKLQDYTGVEKDEVILQCEISK 11002

Qy        302 ---------DG--IAP--NA--------QIFSYKM--------YS-DAGSGFAGDETM-- 329
                       ||  | |  ||        ::: :|:        |: | |:    |:|
Db      11003 ADAPVKWFKDGKEIKPSKNAVIKADGKKRMLILKKALKSDIGQYTCDCGT----DKTSGK 11058

Qy        330 -------------FHAI---------------EDSI------------------------ 337
                          :|::               || |
Db      11059 LDIEDREIKLVRPLHSVEVMETETARFETEISEDDIHANWKLKGEALLQTPDCEIKEEGK 11118

Qy        338 -----KHN--------VDV-------------------------VSVSSG--------- 349
                   ||         ||:                        |:|::|
Db      11119 IHSLVLHNCRLDQTGGVDFQAANVKSSAHLRVKPRVIGLLRPLKDVTVTAGETATFDCEL 11178

Qy        350 ------------------------------------------------------------ 349

Db      11179 SYEDIPVEWYLKGKKLEPSDKVVPRSEGKVHTLTLRDVKLEDAGEVQLTAKDFKTHANLF 11238

Qy        350 -------FT---------------------------------GT-GLVGEKY--------- 360
                    ||                                 ||  |  :||
Db      11239 VKEPPVEFTKPLEDQTVEEGATAVLECEVSRENAKVKWFKNGTEILKSKKYEIVADGRVR 11298

Qy        361 ------------------------------------------------------------ 360

Db      11299 KLVIHDCTPEDIKTYTCDAKDFKTSCNLNVVPPHVEFLRPLTDLQVREKEMARFECELSR 11358

Qy        361 ------W-----------------QAIRAL-------------------RKAGIPMV- 375
                    |                  |:| |                   | :|:  |
```

```
Db      11359 ENAKVKWFKDGAEIKKGKKYDIISKGAVRILVINKCLLDDEAEYSCEVRTARTSGMLTVL 11418

Qy        376 ----VATGNYATSASSSS----------------W---D-------------------- 391
                  | | |:|:   | :                 |   |
Db      11419 EEEAVFTKNLANIEVSETDTIKLVCEVSKPGAEVIWYKGDEEIIETGRYEILTEGRKRIL 11478

Qy        392 LVANNHLKMTDTGNV------TRT----AAHE---------------------------- 413
              :::| ||:  |:||:      :||     :||
Db      11479 VIQNAHLE--DAGNYNCRLPSSRTDGKVKVHELAAEFISKPQNLEILEGEKAEFVCSISK 11536

Qy        414 --------------------DAIA-----------------------VASA--------- 421
                               |:||                        | :|
Db      11537 ESPPVQWKRDDKTLESGDKYDVIADGKKRVLVVKDATLQDMGTYVVMVGAARAAAHLTVI 11596

Qy        422 ---------------KNQTVEFD-KVNIGGESFKY-RNIGAFFDKSK------------- 451
                          : | | |: :||  |   |: ||  | || ||
Db      11597 EKLRIVVPLKDTRVKEQQEVVFNCEVNTEGAKAKWFRNEEAIFDSSKYIILQKDLVYTLR 11656

Qy        452 ------------------------------ITTNED------------------------ 457
                                           |:  ||
Db      11657 IRDAHLDDQANYNVSLTNHRGENVKSAANLIVEEEDLRIVEPLKDIETMEKKSVTFWCKV 11716

Qy        458 ------------GTKAPSKLKFVY-----------------------IGKGQD------- 475
                         | ::|   :: |                       :  |||
Db      11717 NRLNVTLKWTKNGEEVPFDNRVSYRVDKYKHMLTIKDCGFPDEGEYIVTAGQDKSVAELL 11776

Qy        476 ---------------------------------------------------------- 475

Db      11777 IIEAPTEFVEHLEDQTVTEFDDAVFSCQLSREKANVKWYRNGREIKEGKKYKFEKDGSIH 11836

Qy        476 ------------------------------------------QDLI---GLD--------- 482
                                                        ||::   | |
Db      11837 RLIIKDCRLDDECEYACGVEDRKSRARLFVEEIPVEIIRPPQDILEAPGADVVFLAELNK 11896

Qy        483 -------LR----------------GKI---AVMD---------RIYTKD---------- 497
                    ||                |||   :: |         |  :||
Db      11897 DKVEVQWLRNNMVVVQGDKHQMMSEGKIHRLQICDIKPRDQGEYRFIAKDKEARAKLELA 11956

Qy        498 ----LKNA----------------------------FK---------------------- 503
                  :|:|                           ||
Db      11957 AAPKIKTADQDLVVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIDTTAEQTSFRIL 12016

Qy        504 --KAMDKG---------------------------------------------------- 509
                |  |||
Db      12017 EAKKGDKGRYKIVLQNKHGKAEGFINLKVIDVPGPVRNLEVTETFDGEVSLAWEEPLTDG 12076

Qy        510 ---------------------------------------------------------- 509

Db      12077 GSKIIGYVVERRDIKRKTWVLATDRAESCEFTVTGLQKGGVEYLFRVSARNRVGTGEPVE 12136

Qy        510 ------ARA------------IMVVN------------------TVNY------------ 521
                    ||:            |  ||                    :||
Db      12137 TDNPVEARSKYDVPGPPLNVTITDVNRFGVSLTWEPPEYDGGAEITNYVIELRDKTSIRW 12196

Qy        522 ------------------------Y-------NR--------------------------- 524
                                       |       ||
Db      12197 DTAMTVRAEDLSATVTDVVEGQEYSFRVRAQNRIGVGKPSAATPFVKVADPIERPSPPVN 12256

Qy        525 ------------------------------------DNW---------------------- 527
                                                   |||
Db      12257 LTSSDQTQSSVQLKWEPPLKDGGSPILGYIIERCEEGKDNWIRCNMKLVPELTYKVTGLE 12316

Qy        528 --------------------TEL-----------------------------------P 531
                               :|:                                  |
Db      12317 KGNKYLYRVSAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEVIVPNPITILVP 12376

Qy        532 AMGYE-----------------------------------ADEG-----TKSQVFSIS 549
                : ||:                               :|:|      : :| :||
Db      12377 STGYPRPTATWCFGDKVLETGDRVKMKTLSAYAELVISPSERSDKGIYTLKLENRVKTIS 12436

Qy        550 G--------------------------------DDG--------------VKLWNMI 560
                |                               |||              | |:::
Db      12437 GEIDVNVIARPSAPKELKFGDITKDSVHLTWEPPDDDGGSPLTGYVVEKREVSRKTWTKV 12496
```

```
Qy        561 --------------------------N----------------------PDKKTEVK-- 569
                                         |                      ||::  ||
Db      12497 MDFVTDLEFTVPDLVQGKEYLFKVCARNKCGPGEPAYVDEPVNMSTPATVPDPPENVKWR 12556

Qy        570 -RNNKEDF-------------------------KDK-----------------LEQ--YY 584
                 |:    |                       ||                 ||:  :|
Db      12557 DRTANSIFLTWDPPKNDGGSRIKGYIVERCPRGSDKWVACGEPVAETKMEVTGLEEGKWY 12616

Qy        585 -----------------P------IDM----ESF------------------------- 591
                              |      :|     | |
Db      12617 AYRVKALNRQGASKPSRPTEEIQAVDTQEAPEIFLDVKLLAGLTVKAGTKIELPATVTGK 12676

Qy        592 --------------------NSNKP---------------------------------- 596
                                  | |
Db      12677 PEPKITWTKADMILKQDKRITIENVPKKSTVTIVDSKRSDTGTYIIEAVNVCGRATAVVE 12736

Qy        597 -NVGDE--KEIDFKFA-------------P--D--------------TDKELY------- 617
                 || |:  ::  |  :            |  |              || |::
Db      12737 VNVLDKPGPPAAFDITDVTNESCLLTWNPPRDDGGSKITNYVVERRATDSEVWHKLSSTV 12796

Qy        618 -------------KEDI-------------------IV------PAGSTS---------- 629
                          || |                   |:      | |  :
Db      12797 KDTNFKATKLIPNKEYIFRVAAENMYGVGEPVQASPITAKYQFDPPGPPTRLEPSDITKD 12856

Qy        630 --------------------WGPRIDLLLKPD--------------------------- 641
                                  | :|:|    ||
Db      12857 AVTLTWCEPDDDGGSPITGYWVERLD----PDTDKWVRCNKMPVKDTTYRVKGLTNKKKY 12912

Qy        642 --------VSAPGKNIKST-----------------LNVIN----------------GKS 660
                    :: |||  |||                  | |||                | |
Db      12913 RFRVLAENLAGPGKPSKSTEPILIKDPIDPPSPPRWLEVINITKNTADLKWTVPEKDGGS 12972

Qy        661 ---TY----------GY 664
                :|          |:
Db      12973 PITNYIVEKRDVRRKGW 12989
```

```
RESULT 14
ABB66417
ID   ABB66417 standard; Protein; 4345 AA.
XX
AC   ABB66417;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 26043.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10520.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
```

PS Disclosure; SEQ ID NO 26043; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4345 AA;

Query Match 26.5%; Score 1164.4; DB 22; Length 4345;
Best Local Similarity 9.7%; Pred. No. 1.2e-07;
Matches 302; Conservative 159; Mismatches 184; Indels 2456; Gaps 131;

```
Qy          1 YPVVL-------ADTSS----------------------------SEDAL--NISD-KE  21
              | |::        |:||                             : | |  | ||  |
Db         97 YTVAVRTEFAGSGDNSSDLLLKSDLEIFFPKPCEGYLRINDAKLYDTLDELNDNESDSSE 156

Qy         22 KVAE--------NKE--------KHE---------------------------------  31
              |: |        |:|       :|:
Db        157 KTKEYDYYDNLANEESESQNYDNMHPKSSDFNVDLTKNLLRFAFHDGLISEVCPQEQETP 216

Qy         32 ---NI-----------------------------------------------------  33
                ||
Db        217 WVLNIKKGILSAFQNTMMRFDVDANTTETDVSGQCQVQYALEDTDSSYVTIRKTKDINSC 276

Qy         34 ------HSAMETSQ-DFKEKKTA--VIKEKEVVSKNPVIDNNTSNEEAKIKE-------E  77
                  ||:::|:   |:: ||   ::|       | :||||:  |   ||:       :
Db        277 RQRYATHSVLQTTPYTFRDDKTIWPILKSQS--HCNLTIDNNVYKE---IKCLETHLLVP 331

Qy         78 NSNKSQGDYTDS----------------FVNKNTE-----------NP---KKEDKV--  104
              || | | :| |                |:  | |           :|   :  |::
Db        332 FSNASSGALTTSTSRLKLDGVESYSAGEFLEQNPELVERRATLVFDHTPAVKPSHDEIKA 391

Qy        105 ---VYIA-----------EFKD-----------------------------------  112
                 :::            || |
Db        392 ARELLVEMCRVGFPNIQREFIDVFTNFLQTSKSLDYKTLSVLLQRSASTCEQGRNHLLES 451

Qy        113 ----------------------------------------------------KE------  114
                                                                   :|
Db        452 LPFIGSTASYKVMQTLETFHSILEYAKNRLDAEYTLGATAVVHSFCKHHEACEENLRVQQ 511

Qy        115 ----------------SGEKAIKE-----LSSLKNTKV----------------------  131
                             ||:  :|     |  | | |
Db        512 IINLLETEFLNLYNLFKGERRTRERMVILLKGLGNIGVVSSAFAEQLQWIIREDEAPVDI 571

Qy        132 ---------------------------------LYTY-------DRIFNG---SAIE--  145
                                              :|:|       | |  |   | :|
Db        572 RLHGILAFRRVDCARHRSYFLDNYGNYTLNSELRIYSYLQAMRCPDYISVGVIKSILEHE 631

Qy        146 -------------------TTP-------------------------------------  148
                              ::|
Db        632 EINQVGSFVWSHLTNLAKSNSPVRIEAQGLLLNDELSERFKMDIRKFSRNYEHSLFFDEY 691

Qy        149 ---------------------------------------------------------  148

Db        692 NFGTTTDANVIFGTDSYLPRIASVNFTADLFGQSVNFFEFTARAEGLEELAANAFGPKGP 751

Qy        149 ---------------------------------DNL---------------------DK  153
                                                |||                     |
Db        752 LSGQLLRKKLSFLNRWLGNESAEEDDTLENLLSLDNLRLKRKAQEQARASEEEVDEDYDF 811

Qy        154 IKQIEGIS------SVERAQKVQPMM-----------NHARKEIG------------VE-  183
                : :||        |: | |::                 |  | : |           ||
Db        812 EESLEGNKRQKRDVSTTRKQEIDRNVDSLGYKLKYDYNNPRAQFGLRVFGNDLRYFNVES 871

Qy        184 --EAIDYLKSINAPF--------GKNF--------------------------------  200
```

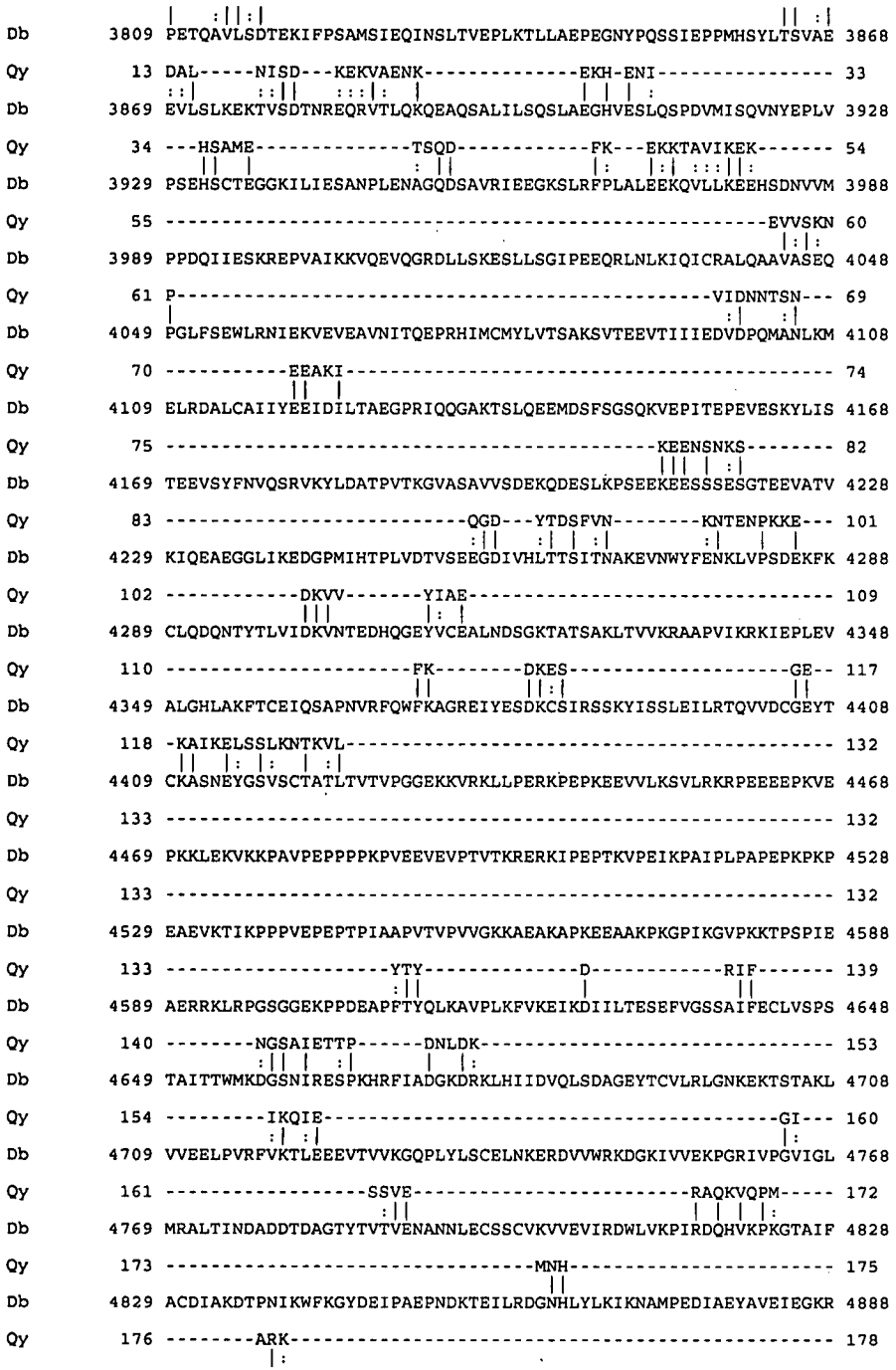
```
                |:: :    | ||        || |
Db        872 LVEVMALAAKFN-PFQQAKNVLSGKEFTYTKSRVFLDASYTVPLAVGLPLAIHAFGASSI 930

Qy        201 DGRGMVISNI---DTGTDYRH---------------KAMRID---------------- 224
              | |  | |:    |  ||: |                :|: |
Db        931 DLR--VSGNLDEMDPPTDW-HFDVEGQFKPSVSVDVITTMQTDMFWEQSGIKVKSNLYSN 987

Qy        225 ---------------------------------------------------------- 224

Db        988 SELVAKLKVRGRNLVSFSFDLPRDKNEIFSVRSELLVQKREEQLPQAGIANRSANSTCTW 1047

Qy        225 -----------------DDAKA--------------SMRFKKEDLKGTDKNY-----W 246
                              | : |              |: :|| ||  : | |     |
Db       1048 PVLDQAIGLQMCSHYSVPDLSNATEIYPSLLLAGPLNFSLILKKSDL--SAKKYVFEYKW 1105

Qy        247 LSD---------------------------KIPHAFN----YYNG-GKITV-EKYD--- 269
                |                          |:| |||    ::||  :::: : ||
Db       1106 --DQQEEDNNFSLVFTTPGSKVPRVLVANVTKVPDAFNASVAFVNGPNRVSAGCSYDGNP 1163

Qy        270 DGR--D-YFDPHG-----------MH---------------IAGI----LAG-------N 289
              | |  | |:|  |           |               | |:    |||       |
Db       1164 DFRRLDIYLDTNGNRSLDLGMELRRHQDFTAWIYNPRMLLAINGVNITGLAGTVKVNEKN 1223

Qy        290 DTEQ-DIK-----------------------------------------NF------NG- 300
                :| |:                                          ||      ||
Db       1224 GIKQHDVDLSFETKKLQAVIKGNVVQSEITTSTNMTIKYRFQANKIEEINFAGKLVNNGD 1283

Qy        301 ------------------------------------I------DGIAPN-----AQIF 311
                                                   |      | : ||      ||
Db       1284 KSKTEYRGNMKLQTSAYPKLNFASESTWLSLQGHTEGMITYNNAPDYVNPNYTSLVRLIF 1343

Qy        312 S---------------------------YK---------------------MYS----- 317
              :                          |:                       :|:
Db       1344 ARSHSEDSFLDGTQTRASLELKLPRSKIDYRILVKHKEHIKNGTEHNVIVGLKYTPEKEI 1403

Qy        318 ----------------DA-------------------------------GSGFAG----- 325
                             ||                               ||:|:|
Db       1404 TGLFSVHLPRRNLFAIDAYMNVTVPEFNSCTASLKVNEKATKDYIIFINGSWFTGHSVAV 1463

Qy        326 -----DET--------------------------------MFHAIE------------ 334
                   | :                              :|  |:
Db       1464 KANYKDRSSRVQALHHLKMIVESPSFNITSLNIIYRRKQLLIFYDIQAKYDQDPYGLTIQ 1523

Qy        335 ---------------------------------------------DSIK--H------NVDV 343
                                                          | |:  |      |||
Db       1524 YASNAHNRNTNAEVRLKVKERDYWINAKLLSEQPKLLQLEIHMDKIRDVHIQVGLLNVDK 1583

Qy        344 ---------------------------------------------VSVSSGFTG-TGLVGE 358
                                                           ::  ||:  ||  |:
Db       1584 RKELSLELKWDANRDPSQRLGLLAEYNSPGTKHYDGNLMITYPERTIHFGFNSFTG--GP 1641

Qy        359 KY-------W---QAIRALRKAGI----------------P-----------MVVATGNY 381
                ||       |   ::|    :|||              |           :::  |:
Db       1642 KYFGKVHASWSINEVIEFEYEAGILPGHTLHNWVKAELRTPFDGWRVNSLDAGIYSLKNL 1701

Qy        382 ATSASSSSW------------DLVANN-------------------------H---LKM 400
                |:  |              |: :|:                        |   :|:
Db       1702 ILVNSTLFWADDQKLQVGYKSDYDVNDQLMSFDVRFGINSTIRDIPTINVKVKHWMDVKK 1761

Qy        401 TDT-------G-NVT----------------R---------------------TAAH--- 412
                :||       | | |              |                       :||
Db       1762 VDTELYLGYSGQNDTFNTYSMDSSWEIEKNQRYNNYSGLVHLVSPFKGYEKGGLVAHFSL 1821

Qy        413 EDAIAVASAKNQTVEFD-------------------KVNI--------------G---- 434
                |: :|: |  ::: ||                  |||               |
Db       1822 SDQRVVSGA--ASLNFDLREFTLTMNGYVKKFTDNMLTVNITTPLEKFGTINARFGLNEK 1879

Qy        435 ---------------G----------------------ESFKYRNIGAFFDKSKI---- 452
                            |                      |||    | |:|:: ::
Db       1880 KRHAVAEVRAPTAALGVEVLADIKNLLNFDVKLSVATPIESFQQAAIFALFNPERVDMRG 1939

Qy        453 ---------------------------------------------------------- 452
```

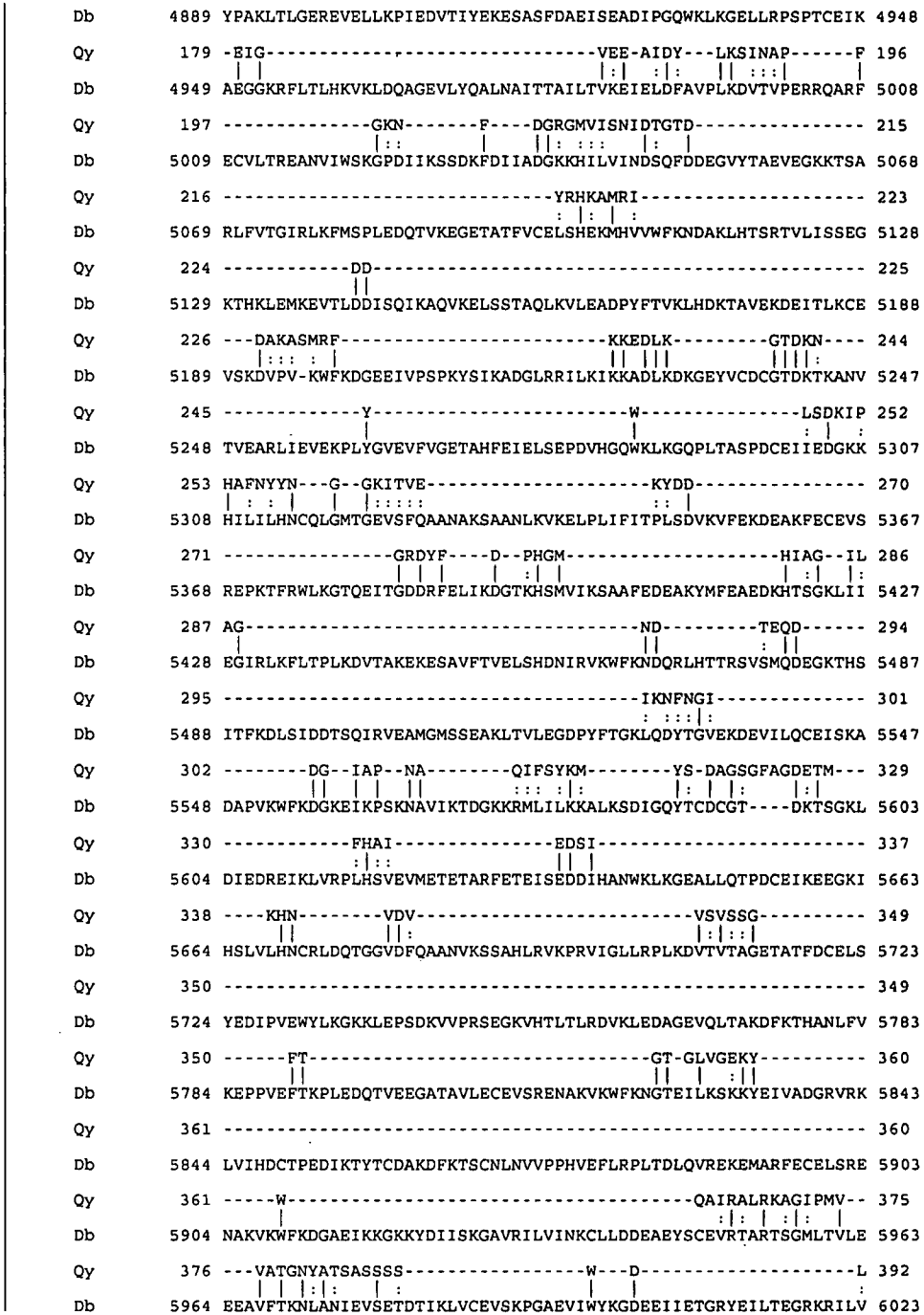
```
Db       1940 LWNNVTLGFTGVWHMQNITDFEYSYHVFTPLAGFEENGFIVQLLKRKEFVFQLHGKMSNY 1999

Qy        453 ------------TTNEDGT---------------KAPSKLK-------FVY------- 469
                          :|: |:                 |:: :|       | |
Db       2000 KLGVKINGEPKSDLVNQLGSNKMELEMLYDADFKPLNAETDYKPDADEEYFSYFTNFQVD 2059

Qy        470 -------IGKGQDQDLIGLDL--------RGKIAVMDRIYTKDL---------------- 498
                    :|   |::| : |         :||: ::||:  :|:
Db       2060 TLVWPTIVGNVDIQEIIDFYLVVGHVELPQGKVEFKDRLHYPDYINVHNLLTVTTPFAVA 2119

Qy        499 --------------KNAF--------------------------------KKAMD----- 507
                           |||                                 :||:|
Db       2120 KNIKSIVEYHVDLNFNAFYERVKFIVNDDKDNTQELGFVFNYTALQDNVKPKAHDVQVTL 2179

Qy        508 ----------------------KG---------------------------------- 509
                                    ||
Db       2180 LTPYEMLHEIYVHGHIELDDNAYKGNISAVTAHTHLSMAASIENEDNFLETSVGIQLETD 2239

Qy        510 ------------------ARAI-------------------------MVVN-------- 517
                              :||                           :||
Db       2240 AIPHYGCQVYFKKDFSAVDKAIDIRFEVTDNGTLNQLHISTDWHTDPSYIVNANGRIKTT 2299

Qy        518 --------------------------------TVNYYNRDN----------WTEL---- 530
                                             :: |  | |           ||::
Db       2300 MLPLQMASTSVLVIQGQNPHLNFDLNLLSQNGQSIAYGARANKKKDVFNIEVWTPMKNFR 2359

Qy        531 ------------------------------------------------------------ 530

Db       2360 NISMHGTAIRSPRDPGRYDVSGFLYRNMATYEVTGAVRMTNSLPIDVVLRVQPKAGGRDG 2419

Qy        531 ----------PAM------------------GY-------EAD-----EGTKSQV----F 546
                         | :                ||       :|     | |: ::    |
Db       2420 VIELNIHEAGPKKIRFSFSAIEDGKMCQMSGGYSVSKTNGAMDFSVLVESTEPEIARINF 2479

Qy        547 ----------SISGD------------DGVKL---------------------------- 556
                        |: ||           | | |
Db       2480 YGNLSPNSEGSLVGDLSLETPWKALGIDTVHLHSDVGFLNKGGHIVGEYKIGQYIGRGSC 2539

Qy        557 ------------------------------------------------------WNM-- 559
                                                                     ||:
Db       2540 LWSWILAEDMQLVLENYLERPNAKPRIVHASAKYQNPGQTFTQLQAGGRLSVDSKWNLDV 2599

Qy        560 ---------------------------------------------INPDKKTE--------- 567
                                                          |||   |
Db       2600 NGSAEYKSVDDFKFRVITALPLPVGDRHQLSASYQGNVISQQFNNPDFVLEASYESFEAQ 2659

Qy        568 ---VKR-------NN------------------KEDFKDKLEQ----------------- 582
                : |       ||                  : ||:   :|
Db       2660 NKLLSRISYKNATNNLKGLGHVEWGKIQNLSVVEGDFELLHKQGAQREFSAKIITPKFKN 2719

Qy        583 ---------Y---------------YP-----IDME-----------SFNSNKP---NVG 599
                       |               ||      |::           |||: |   ||
Db       2720 EHTFALTGSYDLEKSGHHNVVGSLDYPASRRITDLDVSVSSLSNMHGIFNSTLPTFLNV- 2778

Qy        600 DEKEIDFKFA--------------P------------DTD-------------------- 613
                : || |:              |            |:|
Db       2779 SWLKTDFNFTTNNGKSYRYCRCFWPQDTAYFKLNSNYDSDSSNFNHNLNGNVEIEVPLAT 2838

Qy        614 ----------------------------------------KELYKE--DIIV------ 623
                                                      ::|||  || :
Db       2839 RHRADIVYGLQKRRNQDAGNVKVVYNEKQVLDGKYKRLEQAKAPIYKETTDISLENEVKP 2898

Qy        624 -------------PAGS------------------------------------------ 627
                          ||||
Db       2899 LGIHFVSTRDASDPAGSQDVKHIEIYELRNTQNFNLTGELHSRATLKAQDFKVVAIHPNR 2958

Qy        628 -------------------------TSW------------------------------ 630
                                      |:|
Db       2959 AVVLSTKYEDVSPEVVRHHSKLELSETAWIGYNLELGNFSKVGNESQSFALEIFYPKRNL 3018

Qy        631 ---------------------------------------------GPR---------- 633
                                                          | |
Db       3019 SSSGQYYMTDTNFNSDLSFQWLGGNYDQQPKIIHSNLQWKAEPLHRGDREHRTIALTVAH 3078

Qy        634 --------------------------IDLLLKPD--------------------VSAP 645
                                         ||:   ||                    : :
Db       3079 PLLEKDINCKATYYRGLRDLLRTHLTIDYSEYPDQLIELGAQLTDRYSELGHTNYTFHVY 3138

Qy        646 GKNIKSTLNV-ING----------KST----------YG 663
                || | | |:| :||          :||          ||
Db       3139 GKHIASELDVQLNGTLAAMNSYYKTESTAHYKRDIFPARYG 3179
```

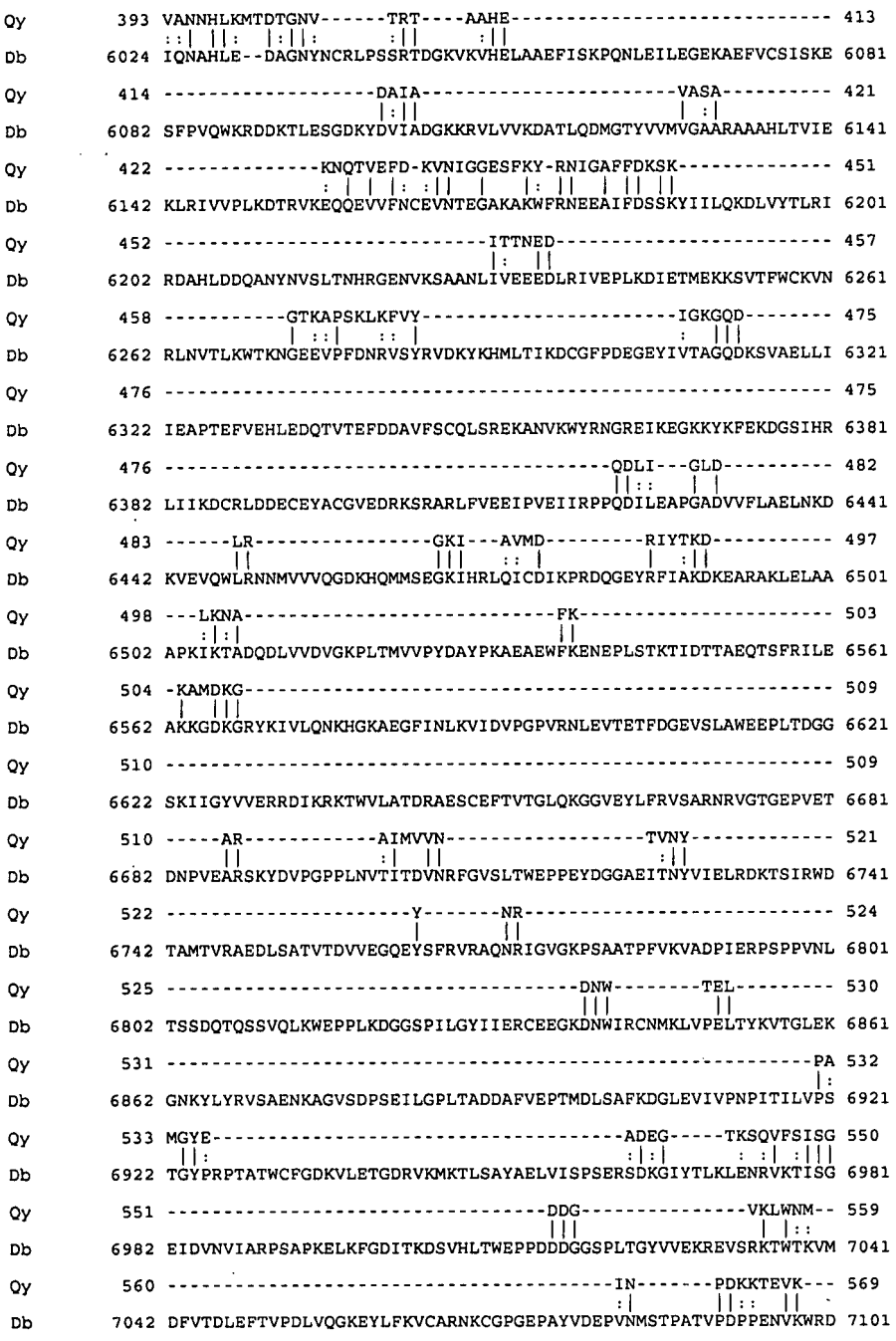
RESULT 15
AAU05396
ID AAU05396 standard; Protein; 26926 AA.
XX
AC AAU05396;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human titin (connectin) protein sequence.
XX
KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW titin-related disease; zebrafish; heart failure; heart disease.
XX
OS Homo sapiens.
XX
PN WO200151666-A1.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-US01212.
XX
PR 12-JAN-2000; 2000US-0175787.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Fishman MC;
XX
DR WPI; 2001-451869/48.
DR N-PSDB; AAS05390.
XX
PT Determining if a subject has or is at risk of developing a
PT titin-related disease or condition, particularly heart failures,
PT comprises detecting the presence of a mutation in the titin gene -
XX
PS Disclosure; Page 57-111; 114pp; English.
XX
CC The present sequence representing human titin (also known as connectin)
CC is described in an invention relating to a novel method for determining
CC whether a subject has or is at risk of developing a titin-related
CC disease or condition. The method comprises analysing a nucleic acid of
CC a sample from the subject and detecting the presence of a mutation
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
CC titin gene, which indicates that the subject has or is at risk of
CC developing a titin-related disease. The zebrafish which has a phenotype
CC similar to mammalian heart failure is used as a model. The method is
CC useful for detecting an increased likelihood of heart disease, such as
CC heart failure, in a patient, so that appropriate intervention can be
CC instituted before any symptoms occur. The method may also be used to
CC facilitate determination of etiology of an existing heart condition,
CC such as heart failure, to identify compounds that can be used to treat
CC or prevent heart conditions, in prenatal genetic screening, e.g. to
CC identify parents who may be carriers of a recessive titin mutation.
CC Compounds identified using the methods may be used to treat patients
CC that have or are at risk of developing heart disease, e.g. heart
CC failure.
XX
SQ Sequence 26926 AA;

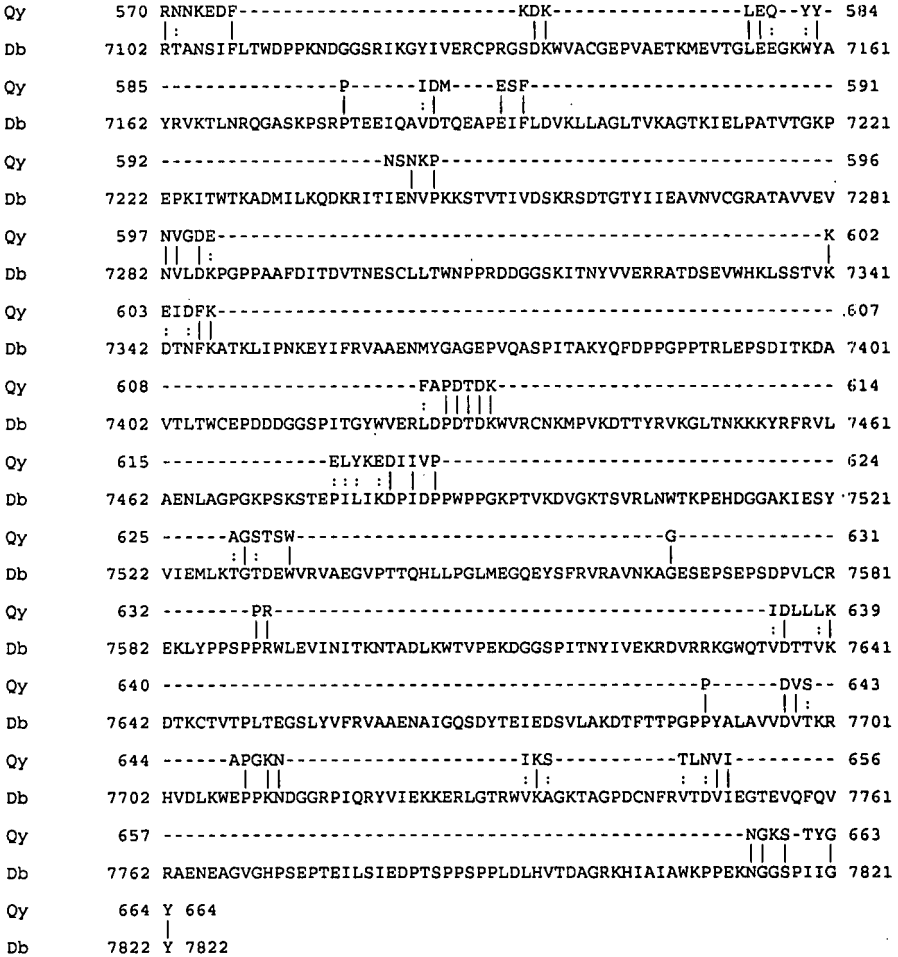
Query Match 26.4%; Score 1161.5; DB 22; Length 26926;
Best Local Similarity 8.0%; Pred. No. 0.00016;
Matches 323; Conservative 169; Mismatches 164; Indels 3365; Gaps 140;

```
Qy          2 P---VVLAD---------------------------------------------TSSSE 12
```









Search completed: October 9, 2003, 15:50:54
Job time : 127 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:55:23 ; Search time 47 Seconds
(without alignments)
2242.435 Million cell updates/sec

Title: US-09-590-991-6
Perfect score: 4398
Sequence: 1 YPVVLADTSSSEDALNISDK.....PGKNIKSTLNVINGKSTYGY 664

Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4398	100.0	664	22	AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21	AAY81710	Streptococcus pneu
3	4395.7	99.9	2140	24	ABU01020	S. pneumoniae type
4	4171	94.8	1007	19	AAW61246	Streptococcus pneu
5	4171	94.8	1007	23	ABP54664	S. pneumoniae SP12
6	3114.7	70.8	26926	22	AAU05396	Human titin (conne
7	3104.2	70.6	31267	24	ABG74786	Human RGS11 protei
8	3098.2	70.4	7107	22	ABB58144	Drosophila melanog
9	3070.2	69.8	5701	22	ABG51536	Human liver peptid
10	3070.2	69.8	5701	22	ABB36684	Peptide #4190 enco
11	3070.2	69.8	5701	22	ABB22021	Protein #4020 enco
12	3070.2	69.8	5701	22	AAM57448	Human brain expres
13	3070.2	69.8	5701	22	AAM69843	Human bone marrow
14	3070.2	69.8	5701	22	AAM05328	Peptide #4010 enco
15	3058.8	69.5	6619	22	ABG23329	Novel human diagno
16	3056.9	69.5	15281	15	AAR44929	T. niveum Cyclospo
17	3032.8	69.0	4643	22	ABB71609	Drosophila melanog
18	3030.6	68.9	7718	22	ABG11811	Novel human diagno
19	3029.6	68.9	4345	22	ABB66417	Drosophila melanog
20	3028.2	68.9	7201	22	ABB71136	Drosophila melanog
21	3026	68.8	4796	22	ABB58665	Drosophila melanog
22	3020.2	68.7	5909	22	ABG23295	Novel human diagno
23	3019.1	68.6	6077	24	ABP95996	White spot syndrom
24	3017.5	68.6	5635	23	ABP60991	Novel human protei
25	3016.9	68.6	4689	22	ABG04530	Novel human diagno
26	3016.6	68.6	4636	22	AAE11937	Human CG168 (or C5
27	3015.1	68.6	6025	22	AAG84939	Shrimp white spot
28	3015	68.6	4655	17	AAR97210	Human kidney calci
29	3015	68.6	4655	17	AAR97211	Human parathyroid
30	3015	68.6	4655	19	AAW43313	Human kidney calci
31	3015	68.6	4655	19	AAW43314	Human parathyroid
32	3013.5	68.5	4599	24	ABP56837	Human LRP1B protei
33	3013	68.5	4655	17	AAR97208	Human calcium sens
34	3011	68.5	4655	19	AAW43311	Human calcium sens
35	3011	68.5	4655	24	ABP56836	Human megalin prot
36	3010.6	68.5	5464	22	ABG11810	Novel human diagno
37	3010	68.4	4655	17	AAR97209	Human placental ca
38	3010	68.4	4655	19	AAW43312	Human placental ca
39	3000.5	68.2	5107	22	ABB65257	Drosophila melanog
40	2999.7	68.2	4472	22	ABB60101	Drosophila melanog
41	2995.6	68.1	4555	23	AAM52106	Rat fat 3 protein
42	2987	67.9	5147	22	ABB59831	Drosophila melanog
43	2979.4	67.7	4741	22	AAU31612	Novel human secret
44	2972.7	67.6	4349	23	AAU79940	Human protocadheri
45	2972.7	67.6	4349	23	ABB97540	Novel human protei

ALIGNMENTS

RESULT 1
AAB48342
ID AAB48342 standard; Protein; 664 AA.
XX
AC AAB48342;
XX
DT 20-APR-2001 (first entry)
XX
DE S. pneumoniae Sp128 polypeptide.
XX
KW Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KW antibacterial; auditory; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO200076540-A2.
XX
PD 21-DEC-2000.
XX
PF 09-JUN-2000; 2000WO-US15925.
XX
PR 10-JUN-1999; 99US-0138453.
XX
PA (MEDI-) MED IMMUNE INC.
XX
PI Adamou JE, Choi GH;
XX
DR WPI; 2001-112197/12.
DR N-PSDB; AAC84741.
XX
PT New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

```
PT   preventing pneumococcal infections, particularly infections caused by
PT   Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT   blood infections  -
XX
PS   Claim 4; Page 47-50; 54pp; English.
XX
CC   The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
CC   from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC   the treatment and prevention of pneumococcal infections, particularly
CC   infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC   bronchial, lung or blood infections. The antigens are used as immunogenic
CC   agents to stimulate an immune response. The antisera and antibodies may
CC   also be used in diagnosing and treating pneumococcal infections.
CC   Recombinant polypeptides serve as a mechanism for stimulating production
CC   of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC   as reagents in other processes such as affinity chromatography. The
CC   present sequence represents the S. pneumoniae Sp128 polypeptide.
XX
SQ   Sequence   664 AA;

  Query Match             100.0%;  Score 4398;  DB 22;  Length 664;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-77;
  Matches  664;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180

Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

```
RESULT 2
AAY81710
ID   AAY81710 standard; Protein; 2120 AA.
XX
AC   AAY81710;
XX
DT   02-JUN-2000  (first entry)
XX
DE   Streptococcus pneumoniae protein sequence ID3.
XX
KW   Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW   bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW   kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW   pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006738-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02452.
XX
PR   27-JUL-1998;   98GB-0016336.
PR   19-MAR-1999;   99US-0125329.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Le Page RWF,  Wells JM,  Hanniffy SB,  Hansbro PM;
XX
DR   WPI; 2000-195301/17.
DR   N-PSDB; AAZ91806.
XX
PT   Streptococcal proteins and polynucleotides useful for diagnosis,
PT   treatment and prophylaxis of bacterial infections  -
XX
PS   Claim 2; Page 41-42; 76pp; English.
XX
CC   This sequence represents a Streptococcus pneumoniae protein of the
CC   invention. The proteins (or their homologues, derivatives and/or
CC   fragments)are useful as immunogens or antigens. Immunogenic or antigenic
CC   compositions comprising the proteins are useful as vaccines and also in
CC   diagnostic assays. The sequences are useful for the detection or
CC   diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC   with them. Agents capable of antagonising, inhibiting or interfering with
CC   the function or expression of the protein or polypeptide are useful in
CC   medical compositions in the treatment or prophylaxis of S. pneumoniae
CC   infection. As the sequences can be used to treat S. pneumoniae infection,
CC   they can be used to treat bacterial pneumonia, which has high rates in
CC   young children, the elderly, and in patients with predisposing conditions
CC   such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC   or with immunosuppressive disorders, especially AIDS. They can also be
CC   used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC   meningitis.
XX
SQ   Sequence   2120 AA;

  Query Match             100.0%;  Score 4398;  DB 21;  Length 2120;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-74;
  Matches  664;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKN 60

Qy         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAI 120

Qy        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEI 180
```

```
Qy        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKEDLKG 240

Qy        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNG 300

Qy        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360

Qy        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 WQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIAVAS 420

Qy        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQDLIG 480

Qy        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEG 540

Qy        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGD 600

Qy        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVINGKS 660

Qy        661 TYGY 664
              ||||
Db        661 TYGY 664
```

```
RESULT 3
ABU01020
ID   ABU01020 standard; Protein; 2140 AA.
XX
AC   ABU01020;
XX
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #590.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media;
KW   ear infection; antiinflammatory; antibacterial; immunostimulant;
KW   auditory; respiratory; gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB02163.
XX
PR   27-MAR-2001; 2001GB-0007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX06302.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT   or ear infection -
XX
PS   Claim 1; SEQ ID No 1180; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the
CC   proteins, treating a patient by administering the protein, DNA or
CC   antibody (in a composition), a kit comprising first and second primers,
CC   which are the nucleic acid cited above or fragments between nucleotides
CC   8-100 of a sequence not defined in the specification, for amplifying a
CC   target sequence contained within a Streptococcus nucleic acid sequence,
CC   where the first primer is substantially complementary to the target
CC   sequence and the second primer is substantially complementary to the
CC   complement of the target sequence, and where the parts of the primers
CC   having substantial complementarity define the termini of the target
CC   sequence to be amplified, assay comprising contacting a test compound
CC   with the protein, and determining whether the test compound binds to the
CC   protein and a Streptococcus  pneumoniae bacterium, where one or more
CC   genes encoding the proteins has been rendered inactive. The proteins,
CC   nucleic acid molecules, antibody and compositions are useful as
CC   medicaments for treating or preventing a disease or infection due to
CC   streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC   sepsis, otitis media or ear infection. They are also useful in developing
CC   vaccines, diagnostics and antibiotics. The methods are useful for
CC   identifying immunodominant proteins. The present sequence is one of
CC   the 2469 proteins expressed by the identified coding regions from the
CC   genomic sequence.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2140 AA;
```

```
  Query Match           99.9%;  Score 4395.7;  DB 24;  Length 2140;
  Best Local Similarity   99.6%;  Pred. No. 5.6e-74;
  Matches  664;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    2;

Qy          1 Y-P--VVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 57
              | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 YAPNEVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVV 78

Qy         58 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 SKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 138

Qy        118 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 KAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMMNHAR 198

Qy        178 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 KEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAKASMRFKKED 258

Qy        238 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 LKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKN 318

Qy        298 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVG 378

Qy        358 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 EKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNHLKMTDTGNVTRTAAHEDAIA 438

Qy        418 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 VASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD 498
```

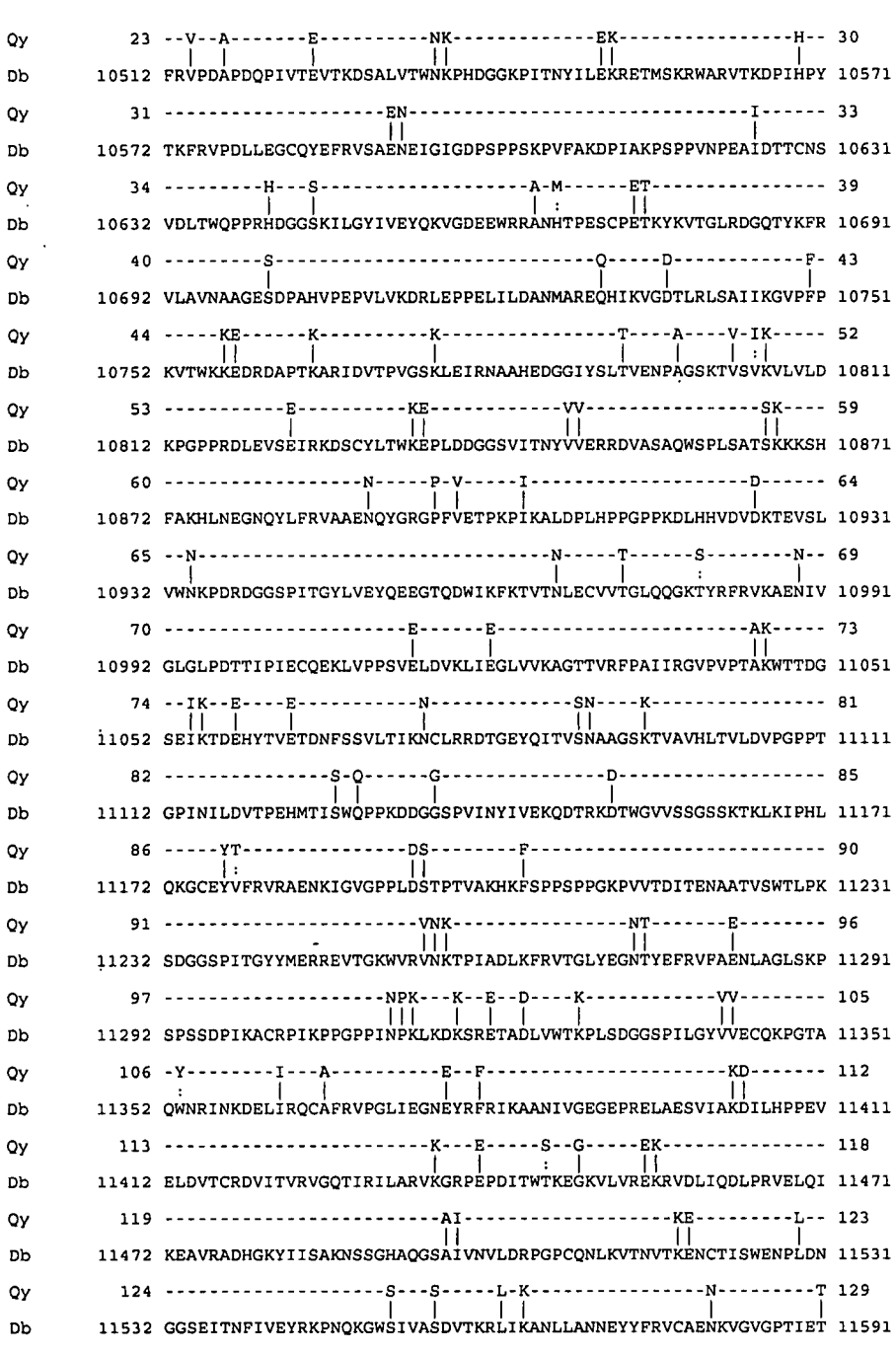
```
Qy        478 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEA 558

Qy        538 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 DEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPN 618

Qy        598 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 VGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVSAPGKNIKSTLNVIN 678

Qy        658 GKSTYGY 664
              |||||||
Db        679 GKSTYGY 685
```

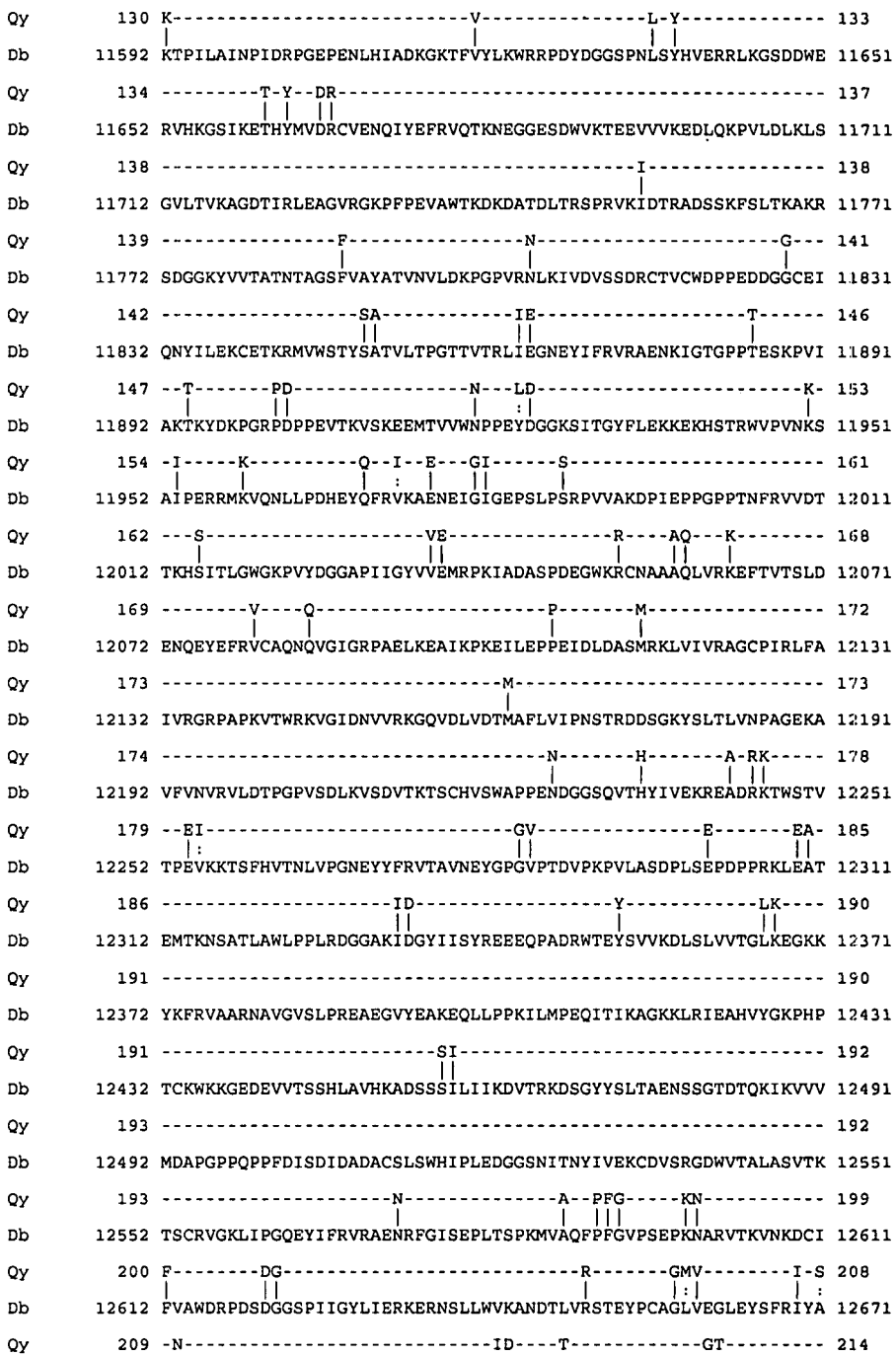
RESULT 4
AAW61246
ID AAW61246 standard; Protein; 1007 AA.
XX
AC AAW61246;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP122 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27431.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 11; Page 92-93; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 1007 AA;

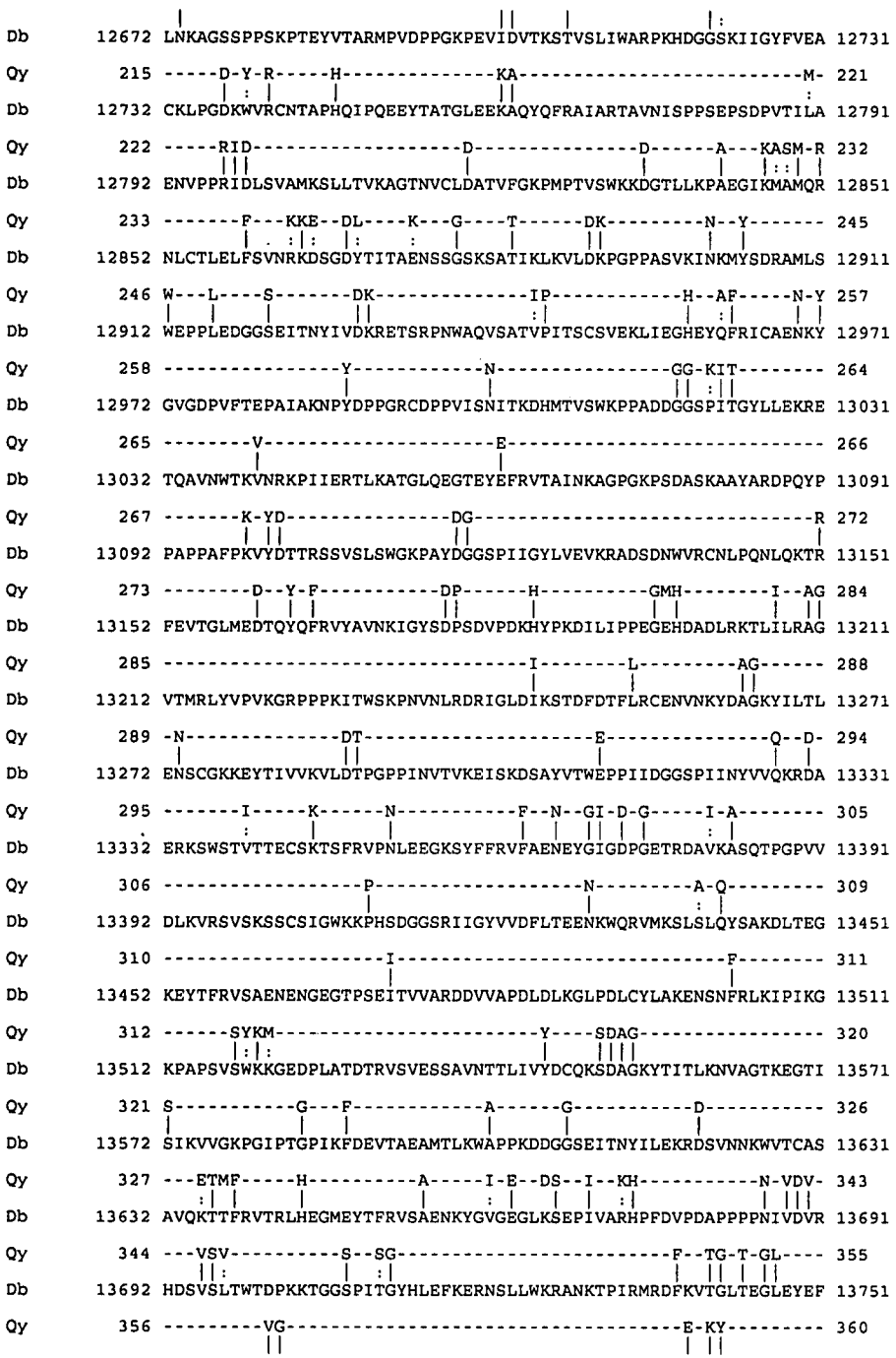
```
  Query Match            94.8%;  Score 4171;  DB 19;  Length 1007;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-71;
  Matches  627;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              |||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627
```

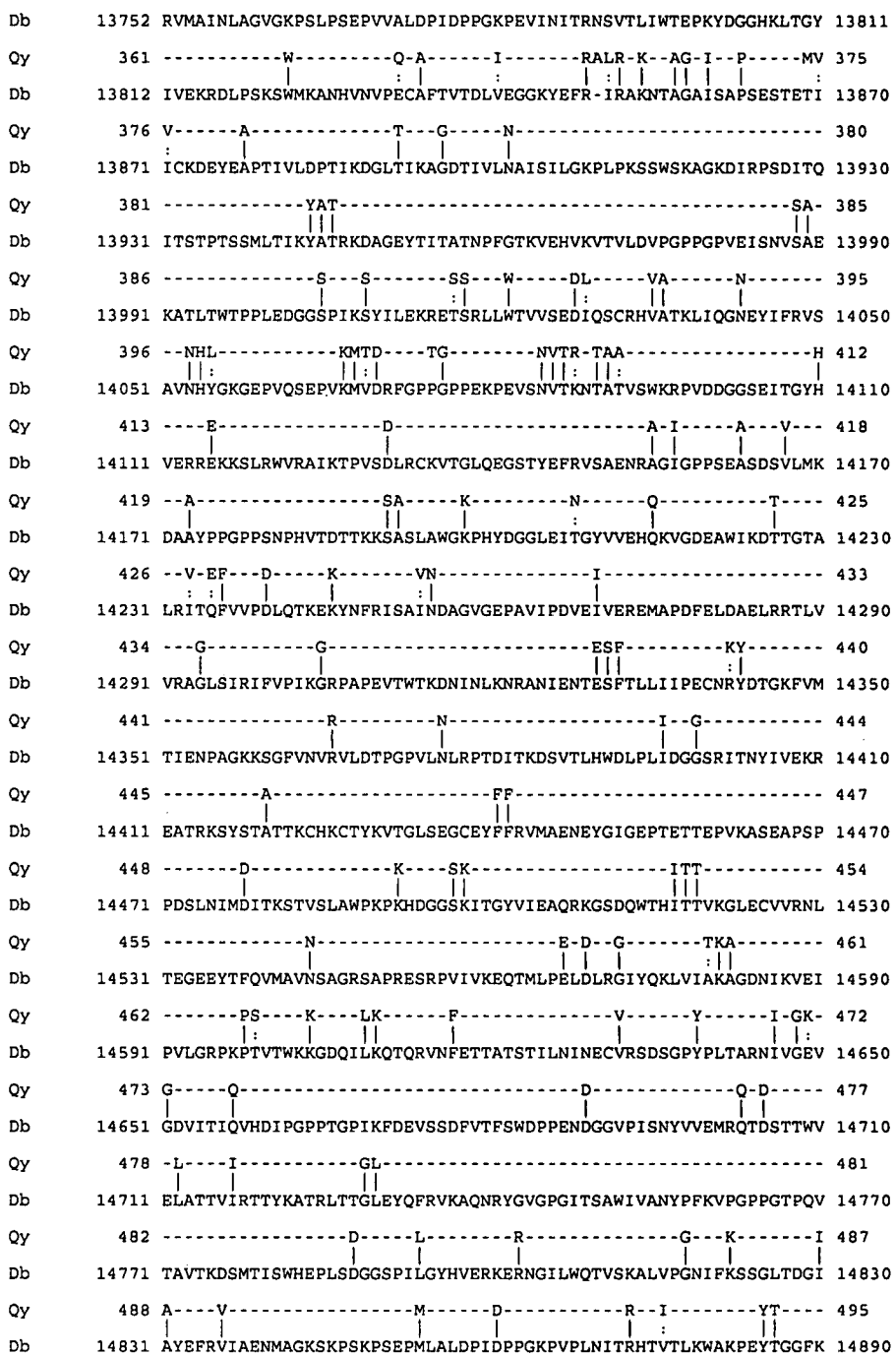
RESULT 5
ABP54664
ID ABP54664 standard; Protein; 1007 AA.
XX
AC ABP54664;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP122 protein sequence SEQ ID NO:216.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.

```
PA   (DOUG/) DOUGHERTY B.
PA   (FANN/) FANNON M R.
PA   (ROSE/) ROSEN C A.
XX
PI   Choi GH,  Kunsch CA,  Barash SC,  Dillon PJ,  Dougherty B,  Fannon MR;
PI   Rosen CA;
XX
DR   WPI; 2002-479261/51.
DR   N-PSDB; ABQ84899.
XX
PT   New Streptococcus pneumoniae antigens, useful for detecting
PT   Streptococcus and for preventing or attenuating disease caused by
PT   Streptococcus infection -
XX
PS   Claim 11; Page 50; 70pp; English.
XX
CC   ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC   Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC   The S. pneumoniae antigens have antibacterial activity and can be
CC   used in vaccines. The S. pneumoniae antigens can also be used to
CC   prevent or attenuate a Streptococcal infection in an animal. The
CC   polynucleotides encoding the S. pneumoniae antigens can be used to
CC   detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC   primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC   which are used in an example from the present invention.
XX
SQ   Sequence   1007 AA;

  Query Match             94.8%;  Score 4171;  DB 23;  Length 1007;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-71;
  Matches  627;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTEN 60

Qy         98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQI 120

Qy        158 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYR 180

Qy        218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDP 240

Qy        278 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300

Qy        338 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360

Qy        398 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNED 420

Qy        458 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVN 480

Qy        518 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFK 540

Qy        578 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLL 600

Qy        638 LKPDVSAPGKNIKSTLNVINGKSTYGY 664
              ||||||||||||||||||||||||||
Db        601 LKPDVSAPGKNIKSTLNVINGKSTYGY 627


RESULT 6
AAU05396
ID   AAU05396 standard; Protein; 26926 AA.
XX
AC   AAU05396;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human titin (connectin) protein sequence.
XX
KW   Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW   titin-related disease; zebrafish; heart failure; heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200151666-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01212.
XX
PR   12-JAN-2000; 2000US-0175787.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman MC;
XX
DR   WPI; 2001-451869/48.
DR   N-PSDB; AAS05390.
XX
PT   Determining if a subject has or is at risk of developing a
PT   titin-related disease or condition, particularly heart failures,
PT   comprises detecting the presence of a mutation in the titin gene  -
XX
PS   Disclosure; Page 57-111; 114pp; English.
XX
CC   The present sequence representing human titin (also known as connectin)
CC   is described in an invention relating to a novel method for determining
CC   whether a subject has or is at risk of developing a titin-related
CC   disease or condition. The method comprises analysing a nucleic acid of
CC   a sample from the subject and detecting the presence of a mutation
CC   (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
CC   titin gene, which indicates that the subject has or is at risk of
CC   developing a titin-related disease. The zebrafish which has a phenotype
CC   similar to mammalian heart failure is used as a model. The method is
CC   useful for detecting an increased likelihood of heart disease, such as
CC   heart failure, in a patient, so that appropriate intervention can be
CC   instituted before any symptoms occur. The method may also be used to
CC   facilitate determination of etiology of an existing heart condition,
CC   such as heart failure, to identify compounds that can be used to treat
CC   or prevent heart conditions, in prenatal genetic screening, e.g. to
CC   identify parents who may be carriers of a recessive titin mutation.
CC   Compounds identified using the methods may be used to treat patients
CC   that have or are at risk of developing heart disease, e.g. heart
CC   failure.
XX
SQ   Sequence   26926 AA;

  Query Match             70.8%;  Score 3114.7;  DB 22;  Length 26926;
  Best Local Similarity   9.5%;  Pred. No. 4.3e-41;
  Matches  583;  Conservative   69;  Mismatches   10;  Indels 5463;  Gaps  479;

Qy          1 Y----------------------------PV----------V------LAD--------- 7
              |                            ||          |      | |
Db      10392 YCVVVENSTGSRKGFCQVNVVDHPGPPVGPVSFDEVTKDYMVISWKPPLDDGGSKITNYI 10451

Qy          8 --------------TS-S-------S---E--D------ALN---ISD-------KEK-- 22
                            || |       |   |  |      | |   |||       | |
Db      10452 IEKKEVGKDVWMPVTSASAKTTCKVSKLLEGKDYIFRIHAENLYGISDPLVSDSMKAKDR 10511
```









```
Qy        496 -------K-DL-----------------------------------KNA----------- 501
                      | ||                                   |||
Db      14891 ITSYIVEKRDLPNGRWLKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPS 14950

Qy        502 -------------------FK-----KA----------------M------------- 506
                              ||     ||                  |
Db      14951 DAITCRDDVEAPKIKVDVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAK 15010

Qy        507 -------------------D---------------------------------------- 507
                              |
Db      15011 LEIKIADFSTNLVNKDSTRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEV 15070

Qy        508 ---K-----------GA-------------R------A-------------------I- 513
                |           ||             |      |                   |
Db      15071 TSEKCVLSWFPPLDDGGAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIF 15130

Qy        514 --M------V-----------VN----------------T-------------------V 519
                |      |           ||                |                   :
Db      15131 RVMAVNKYGVGEPLESEPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEII 15190

Qy        520 NY-------------------------------------------------------Y 522
                ||                                                      |
Db      15191 NYIVERRDKAGQRWIKCNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFY 15250

Qy        523 ------------N-R--D----------N-----------------------W----- 527
                          | |  |          |                       |
Db      15251 KACDTVFKPGPPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTP 15310

Qy        528 -------------TE--------------L--PA----------M------------ 533
                          ||              |  ||          |
Db      15311 PAGLKATSYTITGLTENQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADLR 15370

Qy        534 -------------------------------------------------------G 534
                                                                      |
Db      15371 KVVTIRACCTLRLFVPIKGRPDPEVKWARDHGESLDKASIESASSYTLLIVGNVNRFDSG 15430

Qy        535 -Y----E-------A-------------D----E----------------G------- 540
                |    |       |             |    |                |
Db      15431 KYILTVENSSGSKSAFVNVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIKNYI 15490

Qy        541 --------------T---K-S------Q-----VF-----------------SI--S- 549
                           |   | |      |     :|                  |:  |
Db      15491 VEKRESTRKAYSTVATNCHKTSWKVDQLQEGCSYYFRVLAENEYGIGLPAETAESVKASE 15550

Qy        550 -----G-----D--------------DG--------VKL-------W------------ 557
                  |     |              ||        |::      |
Db      15551 RPLPPGKITLMDVTRNSVSLSWEKPEHDGGSRILGYIVEMQTKGSDKWATCATVKVTEAT 15610

Qy        558 ----------------N-----------------M-I--------N------------ 561
                           |                   : |        |
Db      15611 ITGLIQGEEYSFRVSAQNEKGISDPRQLSVPVIAKDLVIPPAFKLLFNTFTVLAGEDLKV 15670

Qy        562 ---------P------D----K-------K-TE------VK---R-----------N--- 571
                    |      |    |       : ||      :|   |           |
Db      15671 DVPFIGRPTPAVTWHKDNVPLKQTTRVNAESTENNSLLTIKDACREDVGHYVVKLTNSAG 15730

Qy        572 ------N--------------K--E--------------D-------F---K-D----- 578
                    |              |  |              |       :   | |
Db      15731 EAIETLNVIVLDKPGPPTGPVKMDEVTADSITLSWGPPKYDGGGSSINNYIVEKRDTSTTT 15790

Qy        579 ------------K---L----E-Q-------------Y---------YP---------- 585
                         |   |    | |             |         ||
Db      15791 WQIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYPFKVPGPPGTP 15850

Qy        586 -I-----D-ME-----------S----F-------NS------NK---P---------- 596
                :     | ||           |    :       ||      ||   |
Db      15851 VVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNSILWVKLNKTPIPQTKFKTTGLEE 15910

Qy        597 ----------N-VG-------------------------------------------D-- 600
                       | ||                                            |
Db      15911 GVEYEFRVSAENIVGIGKPSKVSECYVARDPCDPPGRPEAIIVTRNSVTLQWKKPTYDGG 15970

Qy        601 --------EK----E----------ID--F----------KF-------A-----P-- 610
                      ||    |          ||  |          :|       |     |
Db      15971 SKITGYIVEKKELPEGRWMKASFTNIIDTHFEVTGLVEDHRYEFRVIARNAAGVFSEPSE 16030

Qy        611 ----------------------DT------------D-----------K---EL--- 616
                               ||                |           |   ||
Db      16031 STGAITARDEVDPPRISMDPKYKDTIVVHAGESFKVDADIYGKPIPTIQWIKGDQELSNT 16090

Qy        617 -------------------------Y----K----E----------D---------II- 622
                                        |    |    |          |         :|
Db      16091 ARLEIKSTDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTVNVKVLDRPGPPEGPVVIS 16150

Qy        623 -V----------P---AGS----------TS---W---------------------- 630
                |          |   ||          ||   |
Db      16151 GVTAEKCTLAWKPPLQDGGSDIINYIVERRETSRLVWTVVDANVQTLSCKVTKLLEGNEY 16210

Qy        631 -----------G---------------------------------------P------- 632
                        |                                       |
Db      16211 TFRIMAVNKYGVGEPLESEPVVAKNPFVVPDAPKAPEVTTVTKDSMIVVWERPASDGGSE 16270

Qy        633 ----------------------------R----I---DL----------L-------- 637
                                           |    |   |:          |
Db      16271 ILGYVLEKRDKEGIRWTRCHKRLIGELRLRVTGLIENHDYEFRVSAENAAGLSEPSPPSA 16330

Qy        638 --------LKP---------D-----V----SAP---------G----K---N------- 648
                      :||         |     |    | |         |    |   |
Db      16331 YQKACDPIYKPGPPNNPKVIDITRSSVFLSWSKPIYDGGCEIQGYIVEKCDVNVGEWTMC 16390

Qy        649 -----I-KST-------L------------N---------------------------- 654
                  | | |       |            |
Db      16391 TPPTGINK-TNIEVEKLLEKHEYNFRICAINKAGVGEHADVPGPIIVEEKLEAPDIDLDL 16449

Qy        655 ----VIN----------------------GK--------------S-T---------Y 662
                   :||                    ||              | |         |
Db      16450 ELRKIINIRAGGSLRLFVPIKGRPTPEVKWGKVDGEIRDAAIIDVTSSFTSLVLDNVNRY 16509

Qy        663 --G-Y 664
                | |
Db      16510 DSGKY 16514
```

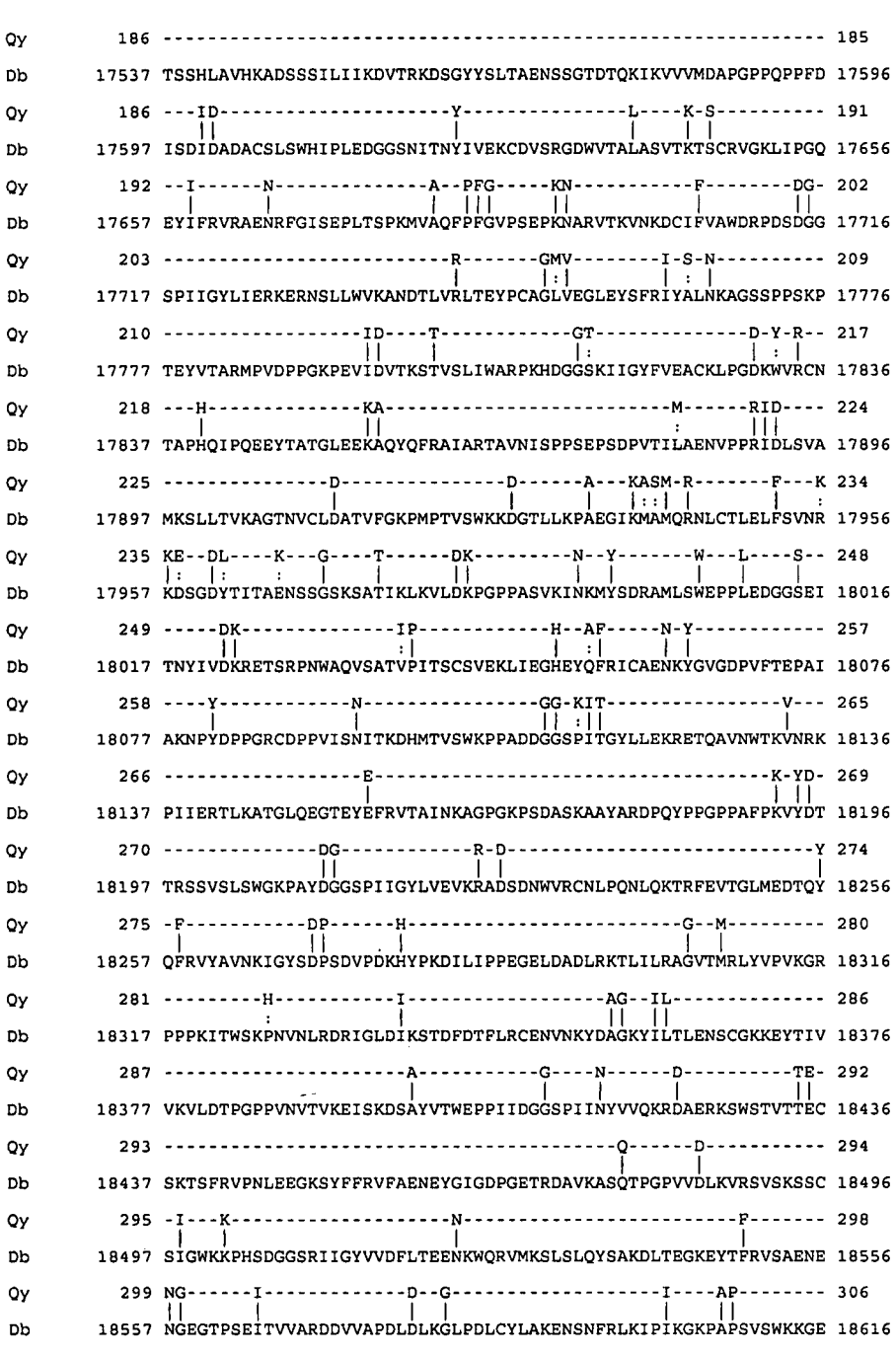
RESULT 7
ABG74786
ID ABG74786 standard; Protein; 31267 AA.
XX
AC ABG74786;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human RGS11 protein.
XX
KW RGS11; human; screening; cardiant; antianginal; gene therapy;
KW heart disorder; cardiac ischaemia; heart failure; angina.
XX
OS Homo sapiens.
XX
PN WO2002103355-A1.
XX
PD 27-DEC-2002.
XX
PF 17-JUN-2002; 2002WO-JP06019.
XX
PR 18-JUN-2001; 2001JP-0183038.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Koyama N, Tanida S, Yamamoto K;
XX
DR WPI; 2003-167557/16.
DR N-PSDB; ABX13540.
XX
PT Screening compounds regulating RGS11 expression and activity for
PT prevention and treatment of heart disease -

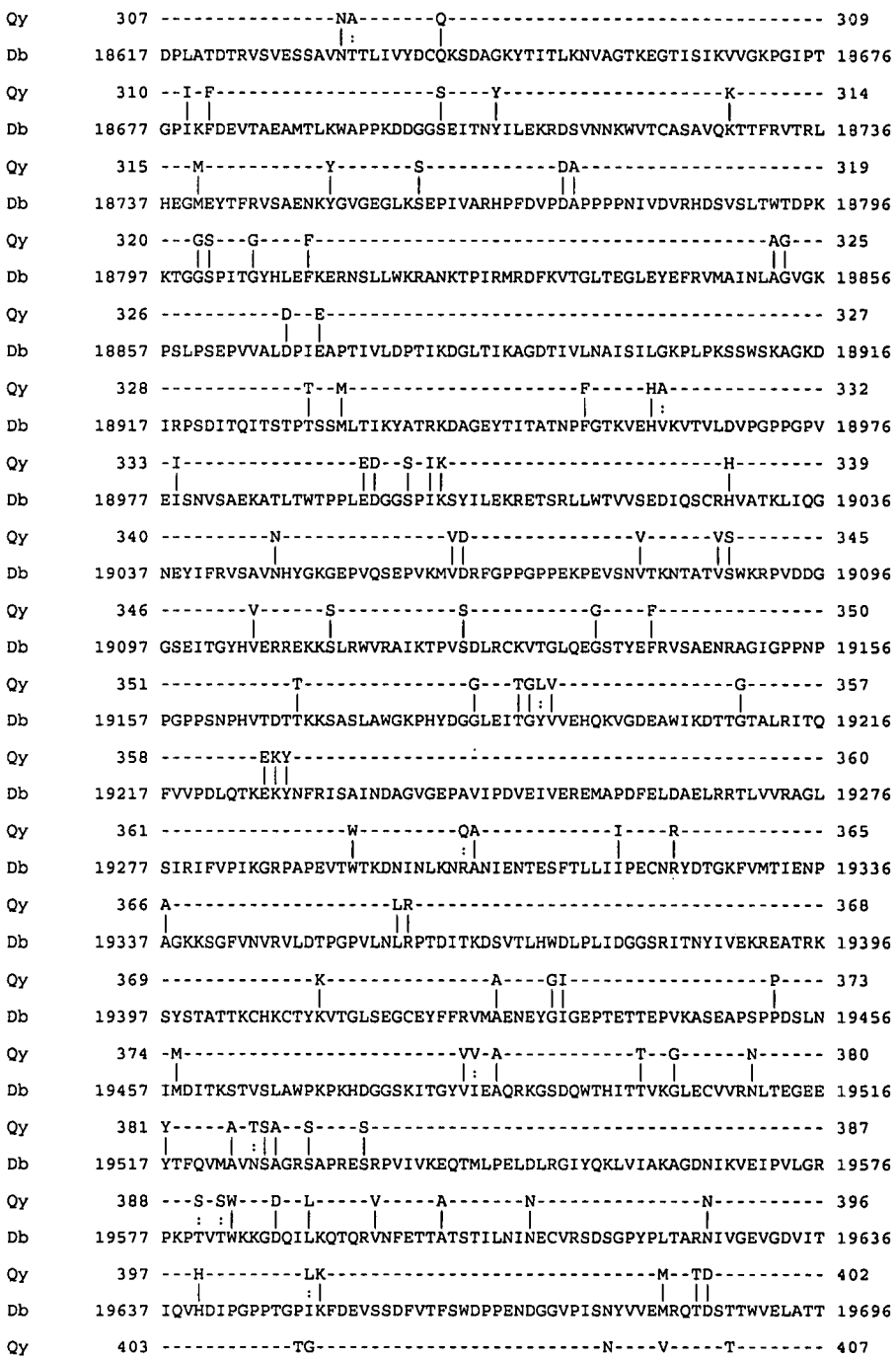
XX
PS Claim 1; Page 59-261; 321pp; Japanese.
XX
CC This invention describes a novel method for screening compounds for their
CC ability to regulate the activity and expression of human RGS11 and its
CC partial peptides and salts, by observing the expression or activity of
CC RGS11 in the presence or absence of the test compound. The products of
CC the invention have cardiant and antianginal activity and can be used for
CC gene therapy. The methods and compositions are useful in the prevention,
CC treatment and diagnosis of heart disorders such as cardiac ischaemia,
CC heart failure and angina. This sequence represents the human RGS11
CC protein described in the disclosure of the invention.
XX
SQ Sequence 31267 AA;

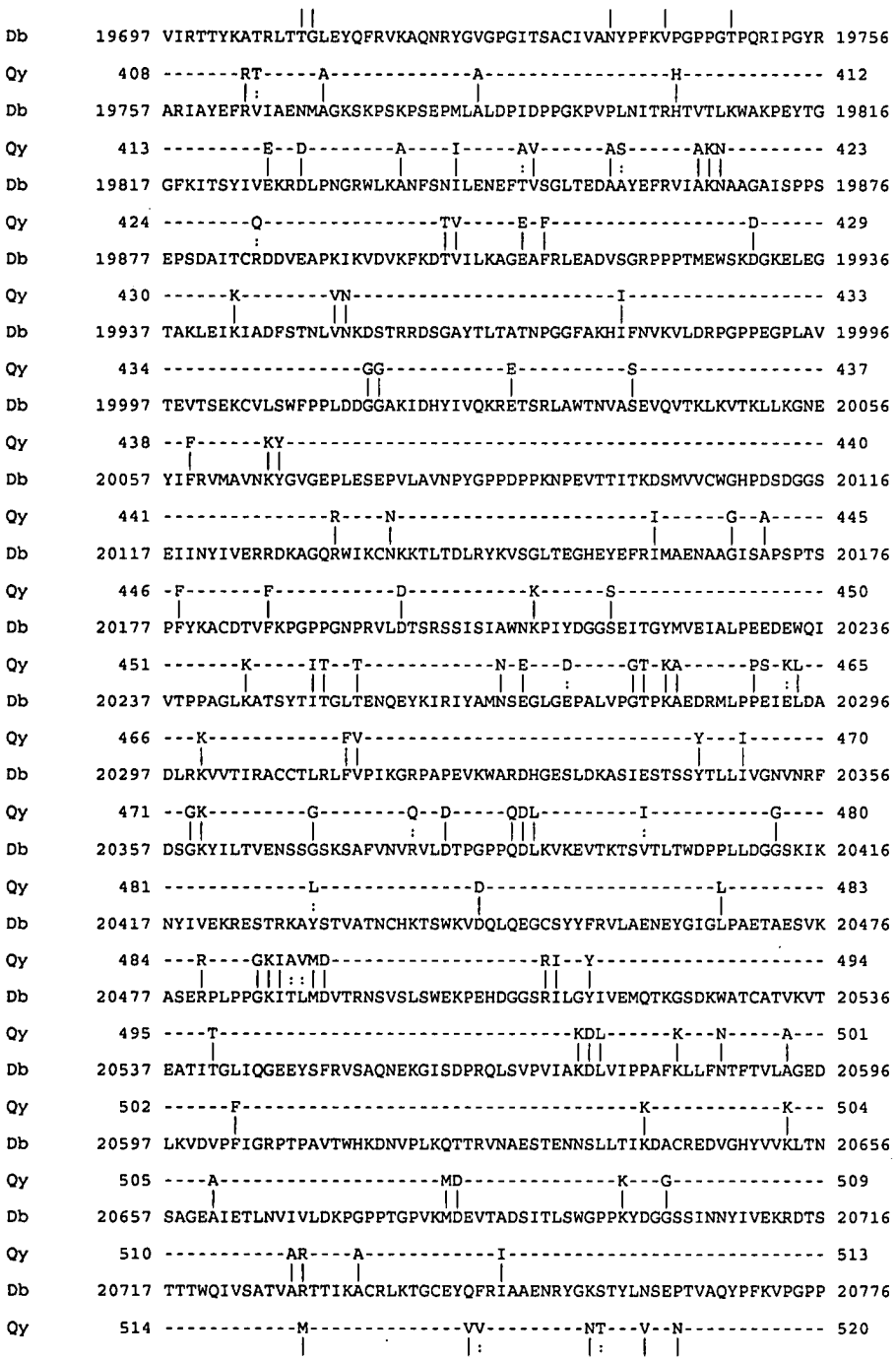
Query Match 70.6%; Score 3104.2; DB 24; Length 31267;
Best Local Similarity 9.4%; Pred. No. 1.6e-40;
Matches 594; Conservative 59; Mismatches 11; Indels 5628; Gaps 490;

```
Qy          1 Y----------------------------PV----------V------LAD--------- 7
              |                            ||          |      | |
Db      15617 YCVVVENSTGSRKGFCQVNVVDHPGPPVGPVSFDEVTKDYMVISWKPPLDDGGSKITNYI 15676

Qy          8 --------------TS-S-------S---E--D------ALN---ISD-------KEK-- 22
                            || |       |   |  |       | |   |||       | |
Db      15677 IEKKEVGKDVWMPVTSASAKTTCKVSKLLEGKDYIFRIHAENLYGISDPLVSDSMKAKDR 15736

Qy         23 --V--A-------E----------NK-------------EK----------------H-- 30
                |  |       |          ||             ||                |
Db      15737 FRVPDAPDQPIVTEVTKDSALVTWNKPHDGGKPITNYILEKRETMSKRWARVTKDPIHPY 15796

Qy         31 --------------------EN-------------------------------I------ 33
                              ||                                 |
Db      15797 TKFRVPDLLEGCQYEFRVSAENEIGIGDPSPPSKPVFAKDPIAKPSPPVNPEAIDTTCNS 15856

Qy         34 ---------H---S--------------------A-M------ET---------------- 39
                      |   |                    | :      ||
Db      15857 VDLTWQPPRHDGGSKILGYIVEYQKVGDEEWRRANHTPESCPETKYKVTGLRDGQTYKFR 15916

Qy         40 ---------S-----------------------------Q-----D------------F- 43
                      |                             |     |            |
Db      15917 VLAVNAAGESDPAHVPEPVLVKDRLEPPELILDANMAREQHIKVGDTLRLSAIIKGVPFP 15976

Qy         44 -----KE------K----------K----------T----A----V-IK----- 52
                  ||      |          |          |    |    | :|
Db      15977 KVTWKKEDRDAPTKARIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLVLD 16036

Qy         53 -----------E----------KE-----------VV----------------SK---- 59
                         |          ||           ||                ||
Db      16037 KPGPPRDLEVSEIRKDSCYLTWKEPLDDGGSVITNYVVERRDVASAQWSPLSATSKKKSH 16096

Qy         60 ------------------N-----P-V-----I--------------------D------ 64
                            |     | |     |                    |
Db      16097 FAKHLNEGNQYLFRVAAENQYGRGPFVETPKPIKALDPLHPPGPPKDLHHVDVDKTEVSL 16156

Qy         65 --N-------------------------------N-----T------S--------N-- 69
                |                               |     |      :        |
Db      16157 VWNKPDRDGGSPITGYLVEYQEEGTQDWIKFKTVTNLECVVTGLQQGKTYRFRVKAENIV 16216

Qy         70 ----------------------E------E-----------------------AK----- 73
                                    |      |                       ||
Db      16217 GLGLPDTTIPIECQEKLVPPSVELDVKLIEGLVVKAGTTVRFPAIIRGVPVPTAKWTTDG 16276

Qy         74 --IK--E----E-----------N-------------SN----K---------------- 81
                 ||  |    |           |             ||    |
Db      16277 SEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNAAGSKTVAVHLTVLDVPGPPT 16336

Qy         82 ---------------S-Q------G--------------D------------------- 85
                           | |      |              |
Db      16337 GPINILDVTPEHMTISWQPPKDDGGSPVINYIVEKQDTRKDTWGVVSSGSSKTKLKIPHL 16396

Qy         86 -----YT--------------DS--------F-------------------------- 90
                  |:              ||        |
```

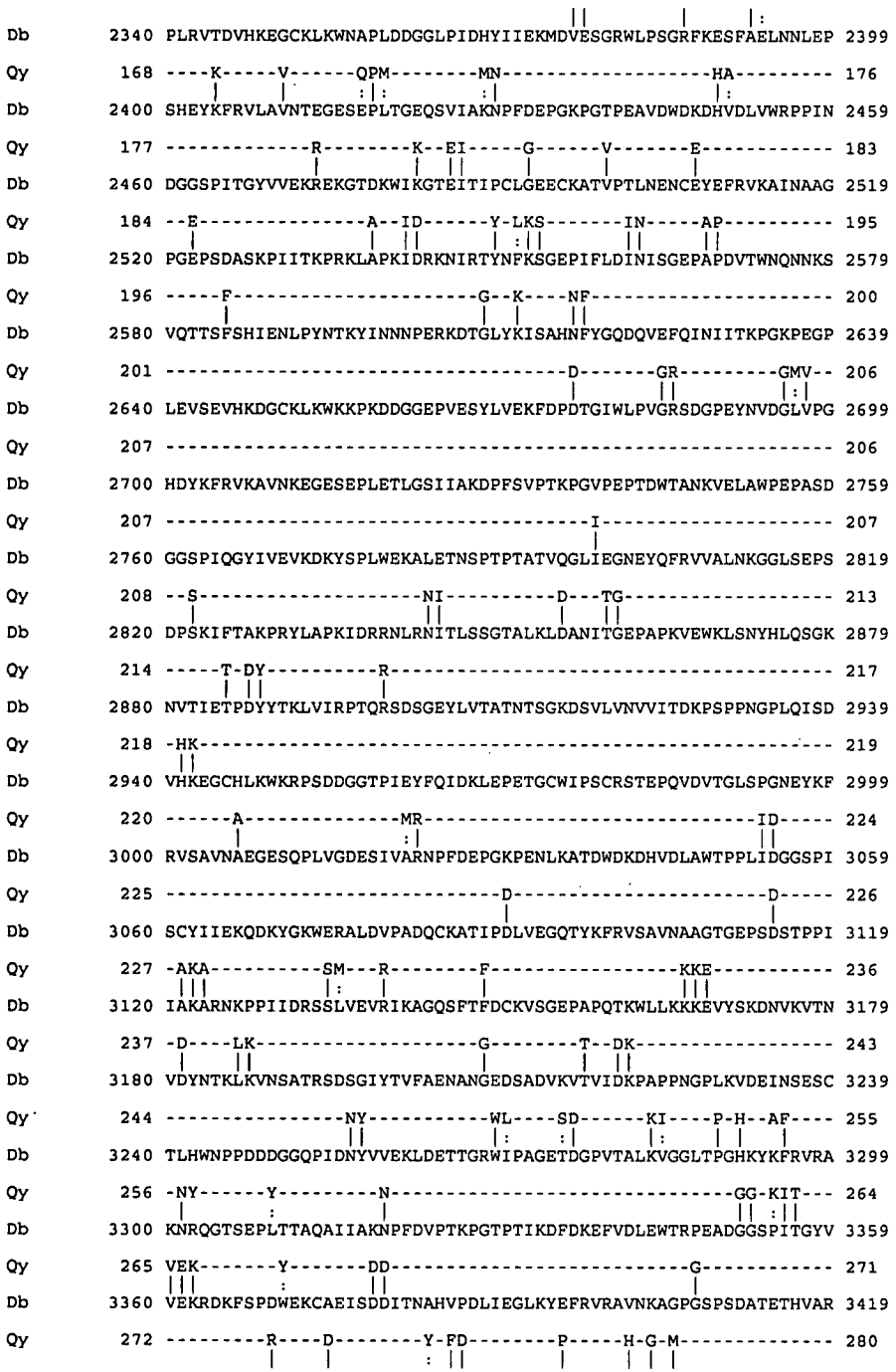
```
Db      16397 QKGCEYVFRVRAENKIGVGPPLDSTPTVAKHKFSPPSPPGKPVVTDITENAATVSWTLPK 16456

Qy         91 ----------------------VNK---------------NT-------E-------- 96
                                    |||               ||       |
Db      16457 SDGGSPITGYYMERREVTGKWVRVNKTPIADLKFRVTGLYEGNTYEFRVFAENLAGLSKP 16516

Qy         97 --------------------NPK---K--E--D----K------------VV-------- 105
                                   |||   |  |  |    |            ||
Db      16517 SPSSDPIKACRPIKPPGPPINPKLKDKSRETADLVWTKPLSDGGSPILGYVVECQKPGTA 16576

Qy        106 -Y--------I---A----------E--F----------------------KD------- 112
               :        |   |          |  |                      ||
Db      16577 QWNRINKDELIRQCAFRVPGLIEGNEYRFRIKAANIVGEGEPRELAESVIAKDILHPPEV 16636

Qy        113 ------------------------K---E-----S--G-----EK--------------- 118
                                       |   |     :  |     ||
Db      16637 ELDVTCRDVITVRVGQTIRILARVKGRPEPDITWTKEGKVLVREKRVDLIQDLPRVELQI 16696

Qy        119 -------------------------AI------------------KE---------L-- 123
                                        ||                  ||         |
Db      16697 KEAVRADHGKYIISAKNSSGHAQGSAIVNVLDRPGPCQNLKVTNVTKENCTISWENPLDN 16756

Qy        124 --------------------S---S-----L-K---------------N---------T 129
                                  |   |     | |               |         |
Db      16757 GGSEITNFIVEYRKPNQKGWSIVASDVTKRLIKANLLANNEYYFRVCAENKVGVGPTIET 16816

Qy        130 K--------------------------V---------------L-Y------------- 133
               |                        |               | |
Db      16817 KTPILAINPIDRPGEPENLHIADKGKTFVYLKWRRPDYDGGSPNLSYHVERRLKGSDDWE 16876

Qy        134 ---------T-Y--DR-----I--F-----N--G------------------------S 142
                      | |  ||     |  |     |  |                        |
Db      16877 RVHKGSIKETHYMVDRCVENQIYEFRVQTKNEGGESDWVKTEEVVVKEDLQKPVLDLKLS 16936

Qy        143 ------A---I----------E---------T--T--P-------------------- 148
                    |   |          |         |  |  |
Db      16937 GVLTVKAGDTIRLEAGVRGKPFPEVAWTKDKDATDLTRSPRVKIDTRADSSKFSLTKAKR 16996

Qy        149 -D--------------------N-LD--------KI---------------------K- 155
                |                    | ||        ||                     :
Db      16997 SDGGKYVVTATNTAGSFVAYATVNVLDKPGPVRNLKIVDVSSDRCTVCWDPPEDDGGCEI 17056

Qy        156 Q--------------------------------IEG------------------------I 160
               |                               |||                        |
Db      17057 QNYILEKCETKRMVWSTYSATVLTPGTTVTRLIEGNEYIFRVRAENKIGTGPPTESKPVI 17116

Qy        161 --------------------S-----------------SV--------E----R------- 165
                                  |                 |:        |    |
Db      17117 AKTKYDKPGRPDPPEVTKVSKEEMTVVWNPPEYDGGKSITGYFLEKKEKHSTRWVPVNKS 17176

Qy        166 A----Q-KVQ----------------------------------------P--------- 171
                |    : |||                                     |
Db      17177 AIPERRMKVQNLLPDHEYQFRVKAENEIGIGEPSLPSRPVVAKDPIEPPGPPTNFRVVDT 17236

Qy        172 --------------------------M---------------------------------- 172
                                        |
Db      17237 TKHSITLGWGKPVYDGGAPIIGYVVEMRPKIADASPDEGWKRCNAAAQLVRKEFTVTSLD 17296

Qy        173 --------M--------------N-------H--------A-RK----------------- 178
                      |              |       |        | ||
Db      17297 ENQEYEFRMSLKHHGHVSWAPPENDGGSQVTHYIVEKREADRKTWSTVTPEVKKTSFHVT 17356

Qy        179 ------------------------------------------------E----------- 179
                                                             |
Db      17357 NLVPGNEYYFRVTAVNEYGPGVPTDVPKPVLASDPLSEPDPPRKLEVTEMTKNSATLAWL 17416

Qy        180 -------------I---------------------------------------------- 180
                          |
Db      17417 PPLRDGGAKIDGYITSYREEEQPADRWTEYSVVKDLSLVVTGLKEGKKYKFRVAARNAVG 17476

Qy        181 --------GV--------------E------------EA-------------------- 185
                      ||              |            ||
Db      17477 VSLPREAEGVYEAKEQLLPPKILMPEQITIKAGKKLRIEAHVYGKPHPTCKWKKGEDEVV 17536
```

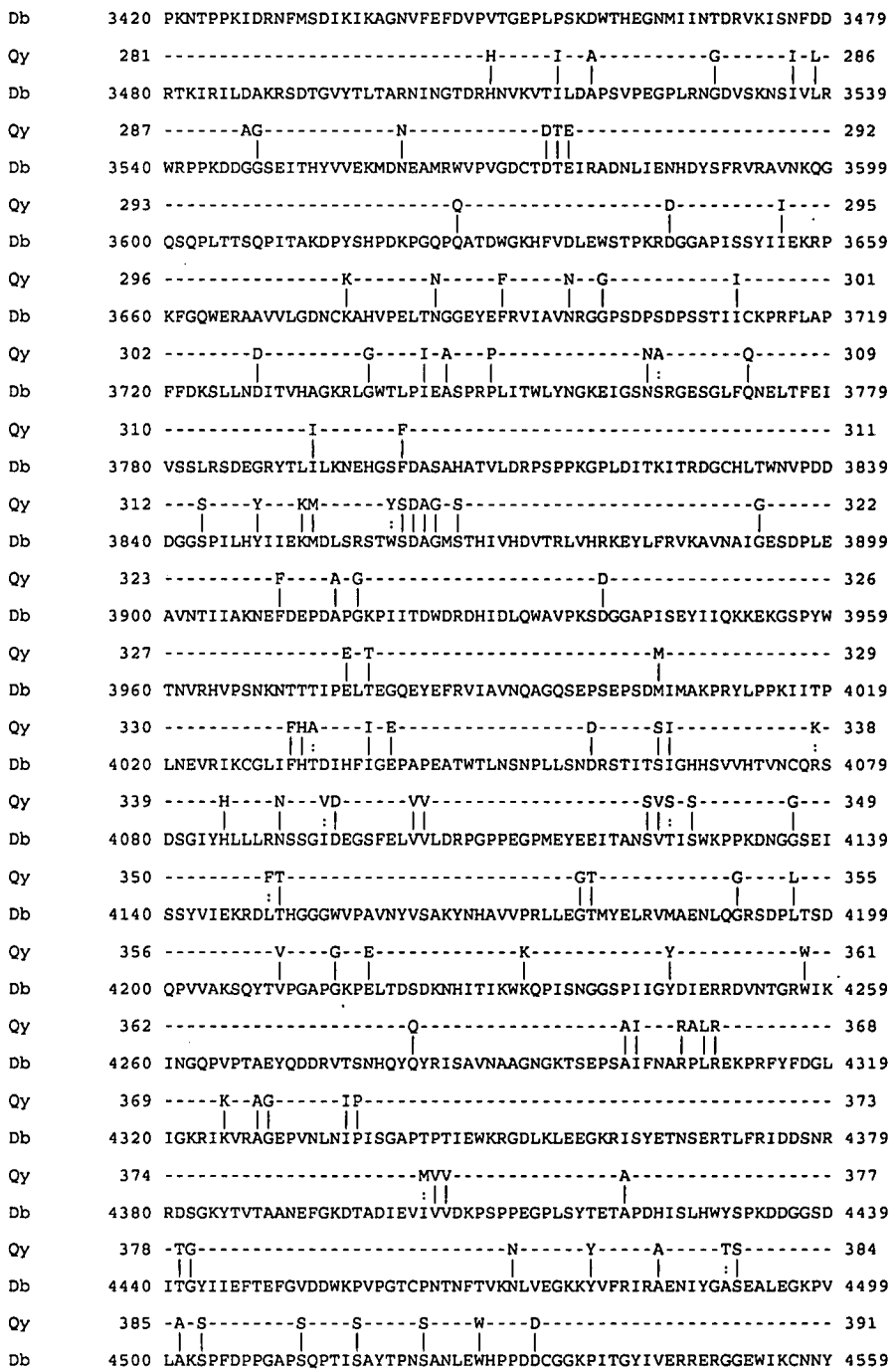


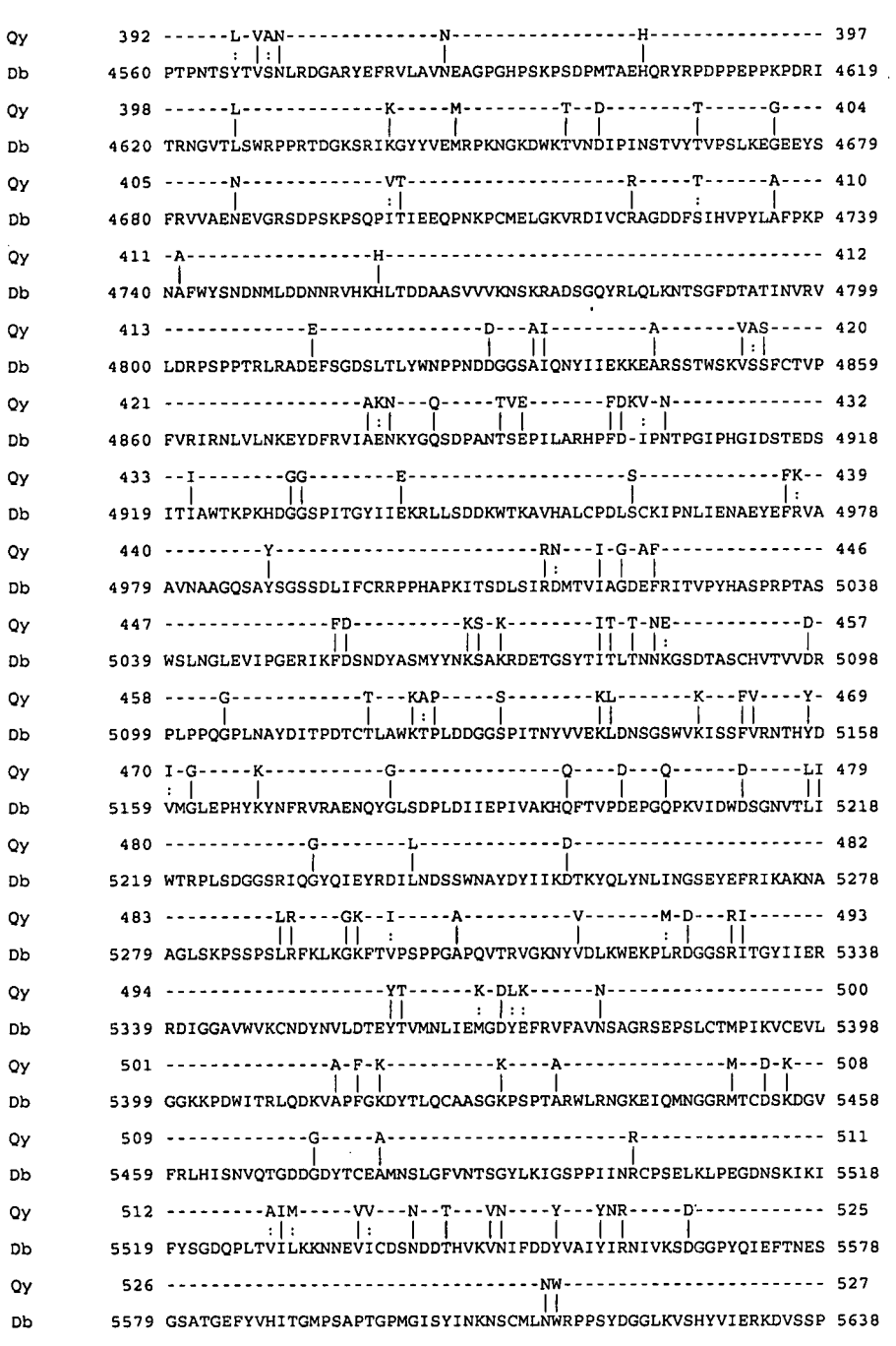


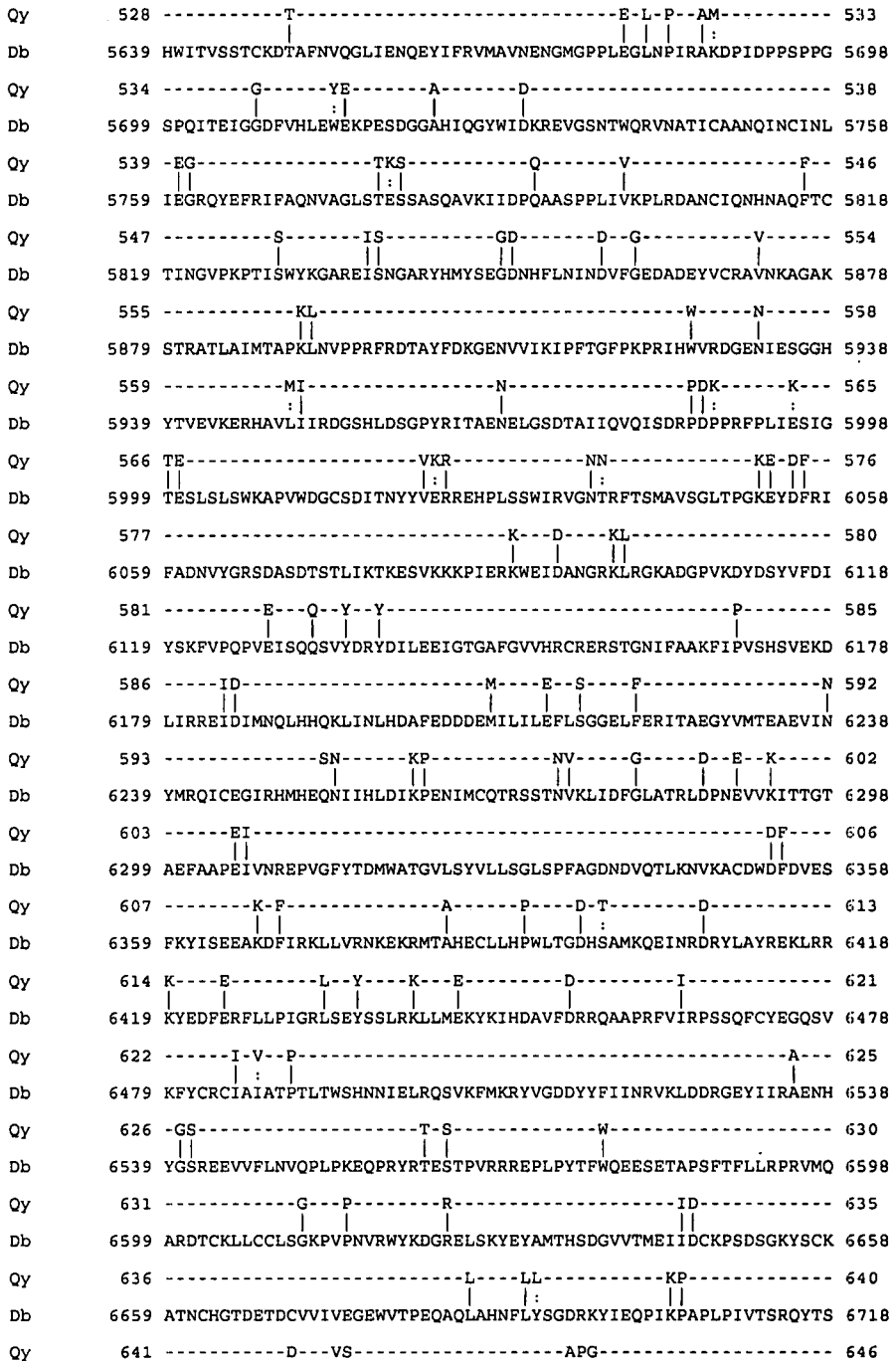

Db 19697 VIRTTYKATRLTTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQRIPGYR 19756
Qy 408 -------RT-----A----------A----------H---------- 412
Db 19757 ARIAYEFRVIAENMAGKSKPSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTG 19816
Qy 413 ---------E--D--------A----I-----AV------AS------AKN--------- 423
Db 19817 GFKITSYIVEKRDLPNGRWLKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPS 19876
Qy 424 --------Q----------TV-----E-F----------D------ 429
Db 19877 EPSDAITCRDDVEAPKIKVDVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEG 19936
Qy 430 ------K--------VN----------I---------- 433
Db 19937 TAKLEIKIADFSTNLVNKDSTRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAV 19996
Qy 434 ----------GG----------E----------S---------- 437
Db 19997 TEVTSEKCVLSWFPPLDDGGAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNE 20056
Qy 438 --F------KY---------- 440
Db 20057 YIFRVMAVNKYGVGEPLESEPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGS 20116
Qy 441 ----------R----N----------I------G--A----- 445
Db 20117 EIINYIVERRDKAGQRWIKCNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTS 20176
Qy 446 -F-------F----------D----------K------S---------- 450
Db 20177 PFYKACDTVFKPGPPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQI 20236
Qy 451 -------K-----IT--T----------N-E---D-----GT-KA------PS-KL-- 465
Db 20237 VTPPAGLKATSYTITGLTENQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDA 20296
Qy 466 ---K----------FV---------- 470
Db 20297 DLRKVVTIRACCTLRLFVPIKGRPAPEVKWARDHGESLDKASIESTSSYTLLIVGNVNRF 20356
Qy 471 --GK---------G--------Q--D-----QDL---------I----------G---- 480
Db 20357 DSGKYILTVENSSGSKSAFVNVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIK 20416
Qy 481 ----------L----------D----------L--------- 483
Db 20417 NYIVEKRESTRKAYSTVATNCHKTSWKVDQLQEGCSYYFRVLAENEYGIGLPAETAESVK 20476
Qy 484 ---R----GKIAVMD----------RI--Y---------- 494
Db 20477 ASERPLPPGKITLMDVTRNSVSLSWEKPEHDGGSRILGYIVEMQTKGSDKWATCATVKVT 20536
Qy 495 ----T----------KDL------K---N-----A--- 501
Db 20537 EATITGLIQGEEYSFRVSAQNEKGISDPRQLSVPVIAKDLVIPPAFKLLFNTFTVLAGED 20596
Qy 502 ------F----------K----------K--- 504
Db 20597 LKVDVPFIGRPTPAVTWHKDNVPLKQTTRVNAESTENNSLLTIKDACREDVGHYVVKLTN 20656
Qy 505 ----A----------MD----------K---G---------- 509
Db 20657 SAGEAIETLNVIVLDKPGPPTGPVKMDEVTADSITLSWGPPKYDGGSSINNYIVEKRDTS 20716
Qy 510 ----------AR----A----------I---------- 513
Db 20717 TTTWQIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYPFKVPGPP 20776
Qy 514 ----------M----------VV---------NT---V--N---------- 520
Db 20777 GTPVVTLSSRDSMEVQWNEPISDGGSRVIGYHLECKERNSILWVKLNKTPIPQTKFKTTG 20836
Qy 521 ------Y-Y-----N----------RD----------N-----W----- 527
Db 20837 LEEGVEYEFRVSAENIVGIGKPSKVSECYVARDPCDPPGRPEAIIVTRNSVTLQWKKPTY 20896
Qy 528 ------T-------ELPA---M----------G------YE---------- 536
Db 20897 DGGSKITGYIVEKKELPEGRWMKASFTNIIDTHFEVTGLVEDHRYEFRVIARNAAGVFSE 20956
Qy 537 ---------A-DE----------G----T----K-SQ-- 544
Db 20957 PSESTGAITARDEVDPPRISMDPKYKDTIVVHAGESFKVDADIYGKPIPTIQWIKGDQEL 21016
Qy 545 -------V----F----SI-------SG----------D--------- 551
Db 21017 SNTARLEIKSTDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTVNVKVLDRPGPPEGPV 21076
Qy 552 ----------DG--------VK------L-W---------- 557
Db 21077 VISGVTAEKCTLAWKPPLQDGGSDIINYIVERRETSRLVWTVVDANVQTLSCKVTKLLEG 21136
Qy 558 N------M-IN----------PD--K--------K---------- 565
Db 21137 NEYTFRIMAVNKYGVGEPLESEPVVAKNPFVVPDAPKAPEVTTVTKDSMIVVWERPASDG 21196
Qy 566 -TE----V----------KR----------N----------N----KE--- 574
Db 21197 GSEILGYVLEKRDKEGIRWTRCHKRLIGELRLRVTGLIENHDYEFRVSAENAAGLSEPSP 21256
Qy 575 --------D----------F----K---D-KLE-Q-Y---------- 583
Db 21257 PSAYQKACDPIYKPGPPNNPKVIDITRSSVFLSWSKPIYDGGCEIQGYIVEKCDVSVGEW 21316
Qy 584 ----------Y----------PI---------- 586
Db 21317 TMCTPPTGINKTNIEVEKLLEKHEYNFRICAINKAGVGEHADVPGPIIVEEKLEAPDIDL 21376
Qy 587 DME----------SF---------N 592
Db 21377 DLELRKIINIRAGGSLRLFVPIKGRPTPEVKWGKVDGEIRDAAIIDVTSSFTSLVLDNVN 21436
Qy 593 ---S---------N----K----------P-N----------V---G-----D---- 600
Db 21437 RYDSGKYTLTLENSSGTKSAFVTVRVLDTPSPPVNLKVTEITKDSVSITWEPPLLDGGSK 21496
Qy 601 ------EK-E----------ID---------F---------- 606
Db 21497 IKNYIVEKREATRKSYAAVVTNCHKNSWKIDQLQEGCSYYFRVTAENEYGIGLPAQTADP 21556
Qy 607 -KFA-----P------D----------T----D----------K--E----- 615
Db 21557 IKVAEVPQPPGKITVDDVTRNSVSLSWTKPEHDGGSKIIQYIVEMQAKHSEKWSECARVK 21616
Qy 616 --------L-----Y-------KE----D-------I------I---V-PA--------G 626
Db 21617 SLQAVITNLTQGEEYLFRVVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVG 21676
Qy 627 ---------S-----T-SW---G-P-----RI------DL--L---------- 637
Db 21677 QDLKIEVPISGRPKPTITWTKDGLPLKQTTRINVTDSLDLTTLSIKETHKDDGGQYGITV 21736
Qy 638 ----------L-KP----------DVSA---------P----G---KN-I--K-- 650
Db 21737 ANVVGQKTASIEIVTLDKPDPPKGPVKFDDVSAESITLSWNPPLYTGGCQITNYIVQKRD 21796
Qy 651 ---------S------TL----------N----------VI---------- 656
Db 21797 TTTTVWDVVSATVARTTLKVTKLKTGTEYQFRIFAENRYGQSFALESDPIVAQYPYKEPG 21856
Qy 657 -------N----------G-------K----S---T---Y-------GY 664
Db 21857 PPGKASKNSECYVARDPCDPPGTPEPIMVKRNEITLQWTKPVYDGGSMITGY 21908

```
RESULT 8
ABB58144
ID   ABB58144 standard; Protein; 7107 AA.
XX
AC   ABB58144;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 1224.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02247.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   7107 AA;

  Query Match             70.4%;  Score 3098.2;  DB 22;  Length 7107;
  Best Local Similarity   9.7%;  Pred. No. 6.1e-44;
  Matches  579;  Conservative   74;  Mismatches   10;  Indels 5278;  Gaps  475;

Qy          1 Y-------------------P-VV-----------------L--A------D------- 7
              |                   | :|                 |  |      |
Db       1020 YQFRVIAVNKAGLSPPSDASVPQIVKYKKLKPRIDRSNLKPLLIRAGKPIRYDVNVRGEP 1079

Qy          8 ----T-------------SSSE--------------------------------D-A- 14
                  |             ||||                                | |
Db       1080 APVITWYQNDKELKPEELPSSSEIKNIPYNTKISIIETVRKHTGIYKIIAVNEHGQDEAT 1139

Qy         15 --LNI----S-------------D-------K-E----K-V-A---E------------ 25
                :||    |             |       | |    | : |   |
Db       1140 VEVNILAPPSKPRGPLDVKDVTKDSCKLKWKKPEDDGGKPISAYQVEKFDKKQGRWVPLG 1199

Qy         26 ----N------K---EKHE---------------------------------------NI- 33
                |      |   | ||                                       ||
Db       1200 RTSANDTEFDVKGLQEGHEYQFRVKAINEEGESDPLDSDDSIIAKNPYDAASKPGTPNIV 1259

Qy         34 ----H----------S---A----------M------E-----TS------------- 40
                  |          |   |          :      |     ||
Db       1260 DYNEHMVKLKWEAPRSDGGAPISGYIIEKKDKFSPIWDEILSTNTSVPEATVEGLVEGNI 1319

Qy         41 -Q-----------------------------------D-------------F----K-E- 45
               |                                   |             |    | |
Db       1320 YQFRVRAVNKAGFSDPSDATEPHLAKPRNLKPYINRDKMKPIKVRAGQPVKFDVDVKGEP 1379

Qy         46 --------K-----------------K-----T-------------------A---- 49
                     |                 |     |                     |
Db       1380 APSLTWFLKETELTSTGQVRLENIDYNTKLTLLDTDRKQSGQYKLRAENINGVDEAVVEV 1439

Qy         50 ------------------------------------------VI---------K-------- 52
                                                        ||         |
Db       1440 IILDKPSKPEGPIEVSDIHKEGCKLKWRKPKDDGGIPITGYVIEKMDTATGKWVPAGSVD 1499

Qy         53 -EK---E-------------V--V------------S----KNP---------------- 61
               ||   |             |  |            |    |||
Db       1500 PEKYDIEIKGLDPNHRYQFRVKAVNEEGESEPLETESAITAKNPFDVSAPPGLPELEDWD 1559

Qy         62 ---V--------I-D-----NN-----------------T-S------------------ 68
                 |        | |     :|                 | |
Db       1560 EHHVKLKWEPPIRDGGSPITNYIIEVMDKDSGEFVKAVETDSPVCKGVVKKLEEGQQYKF 1619

Qy         69 -----N---------EE-----AK----------------I-K-------------E---- 76
                   |         |:     ||                | |             |
Db       1620 RVRAVNKAGPSDPSEQTNWHVAKPRFLKPHIDRVNLKPVIVKTGLSISLDINIRGEPAPK 1679

Qy         77 -E----NS----------------N-K--------SQ----------------------- 83
               |    ||                | |        ||
Db       1680 VEWFFNNSSVTSDEHSVKIDNVDYNTKFFVMRAQRSQSGKYIIKATNEVGEDEAELEVTV 1739

Qy         84 --------------------------------G---DY-----TD-----------S--- 89
                                             |   ||      |           |
Db       1740 LGKPGKPKGPLQVNDITKHSCKLKWEKPDDDGGSPIDYYEIEKLDPHTGQWLPCGKSTEP 1799

Qy         90 --------------F----VNK-----N--TE------NP--------KKE----DK--V 104
                           |    |||     :  ||      ||        |:|    ||  |
Db       1800 EAKVIGLHEGKAYKFRVRAVNKEGESEDLETEKPIIAKNPYDEPDRPGKPEPTNWDKDFV 1859

Qy        105 --------------V--Y-I------------A--------------------EP---- 110
                           :  | |            |                    ||
Db       1860 DLAWDPPKNDGGAPIQKYVIQMRDKSGRAWVDSATVPGDKCNGTVTGVEEGHEYEFRIVA 1919

Qy        111 ----------------------K---D-----KE---SGEK----A-IK---------- 121
                                   |   |     |:   ||::    | ||
Db       1920 VNKAGPSDPSDVSKSVIAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKAEPPAKVTWT 1979

Qy        122 ----EL-SS--LK--N-----T----KV------LY----------------------- 133
                  |: :|  :|  |     |    ||      :|
Db       1980 YNKTEIKTSDHIKIENEDYKTTFIMPKVKRADRGIYIVTAKNDSGSDTVEVELEVLCKPS 2039

Qy        134 --------------T-----------------Y----------------------D-- 136
                           |                 |                      |
Db       2040 KPKGPLAVSNVTAETLHLKWEKPEDDGGDPIEQYLVERMDTETGRWVPVLTTKTPEADVT 2099

Qy        137 ----------R----------------I-------F----------------N------- 140
                       |                |       |                |
Db       2100 GLTEGKEYLFRVKAVNSEGESEPLVTDIPTKAKNPFDAADTPGKPQIVDWSGNHCDLKWR 2159

Qy        141 ------G-S-----------------AIET-TPD-----N-L----------------- 151
                   | |                 |:|| |||     | |
Db       2160 APEDDGGASITGYIVERKDPNTGKWQKALETSTPDCKARVNDLIAGNKYQFRIMAVNKAG 2219

Qy        152 ---------------------D----K---IK--Q-----IE--G-------------- 159
                                  |    |   ||  |     |:  |
Db       2220 KSKPSEPSDQMTAKDRFAPPKIDRTNIKDITIKAGQHIRFDIKVSGEPPATKVWLHNKAR 2279

Qy        160 ----------------------IS--------------------S-------------- 162
                                   ||                    |
Db       2280 LENDDSNYNIDMESYRTKLTVPISKRFHSGKYTLKAENESGRDEASFEVIVLDKPGPPEG 2339

Qy        163 ------------------------------------VE--------R-----AQ------ 167
```



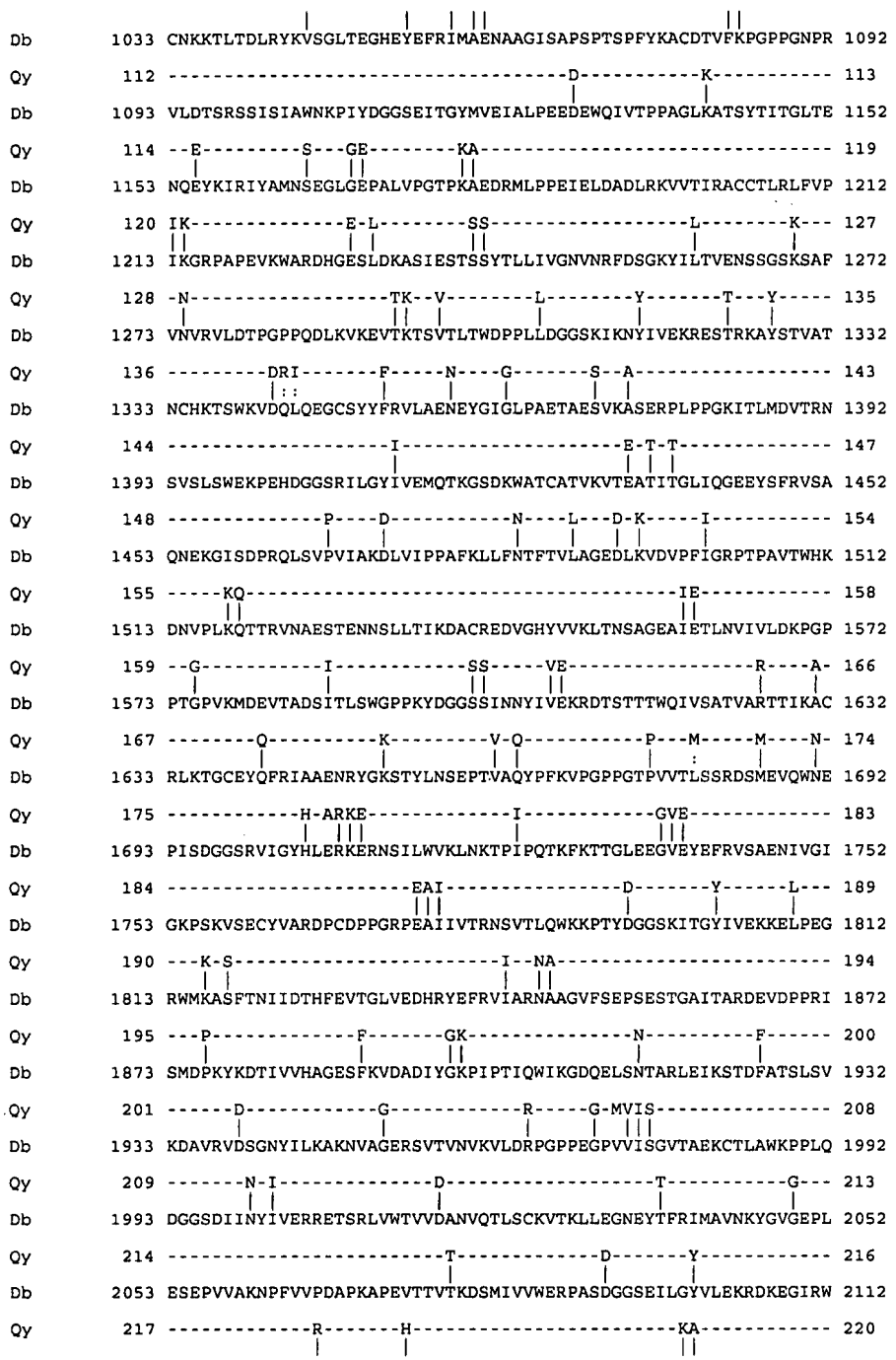


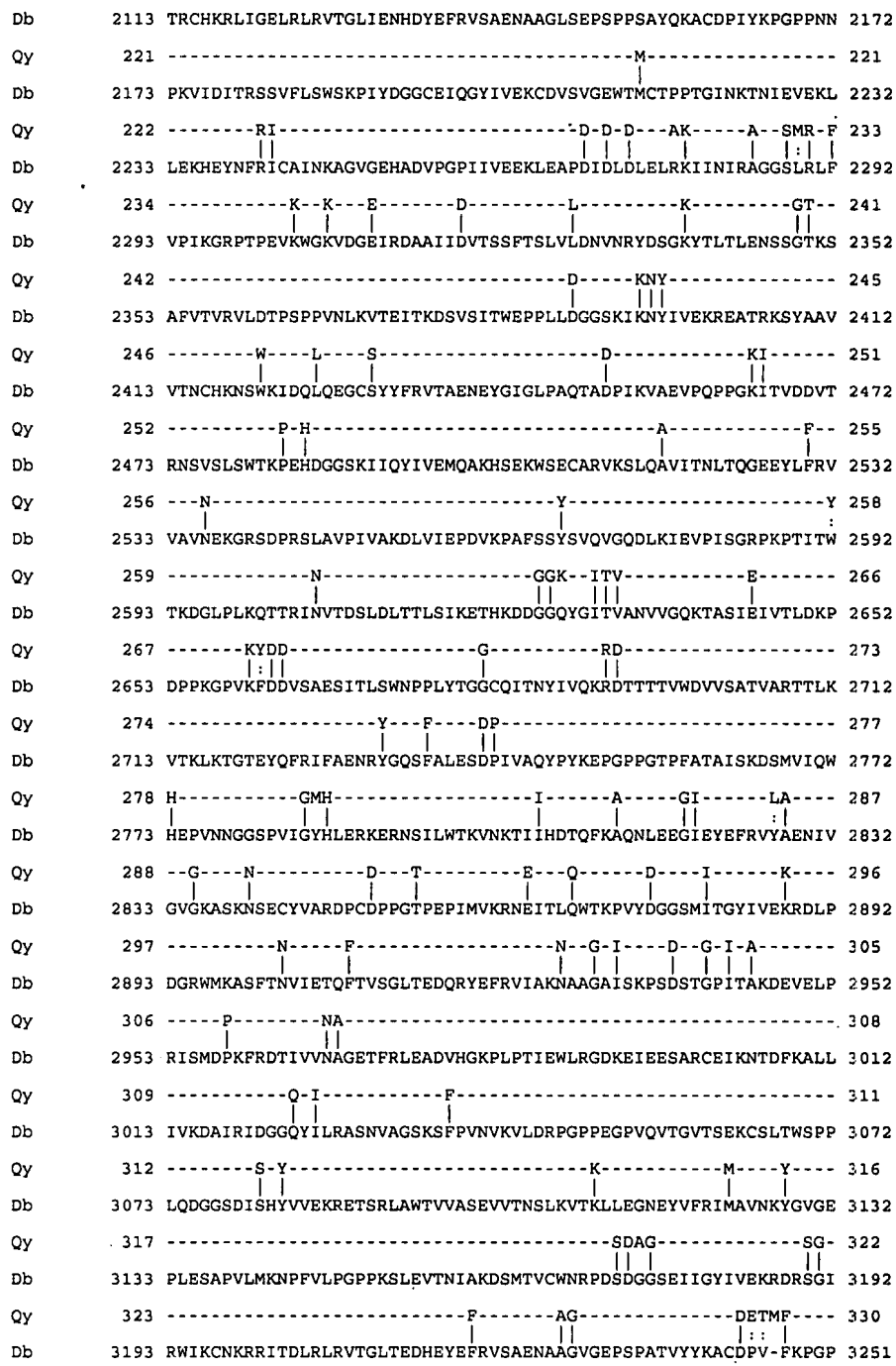


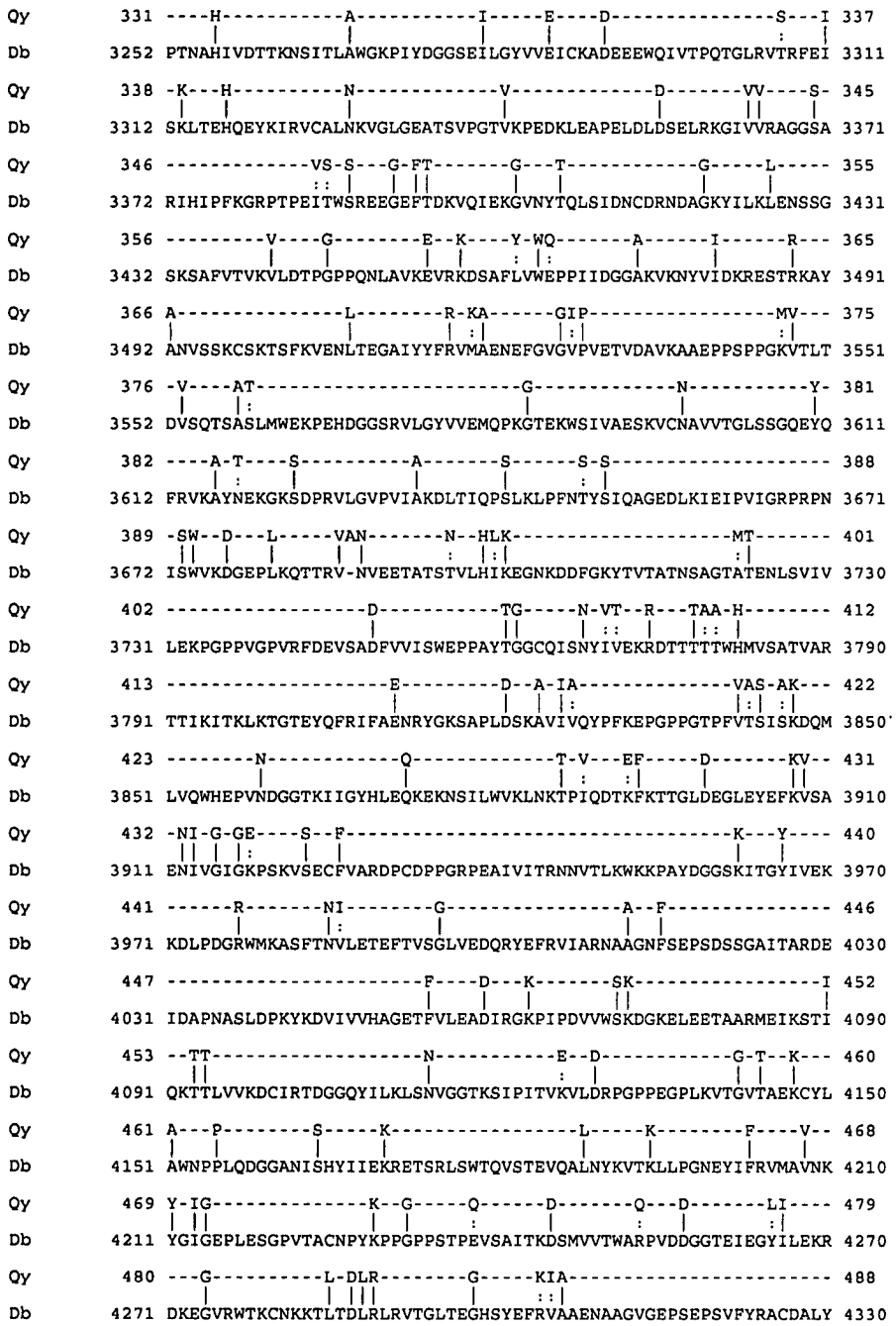


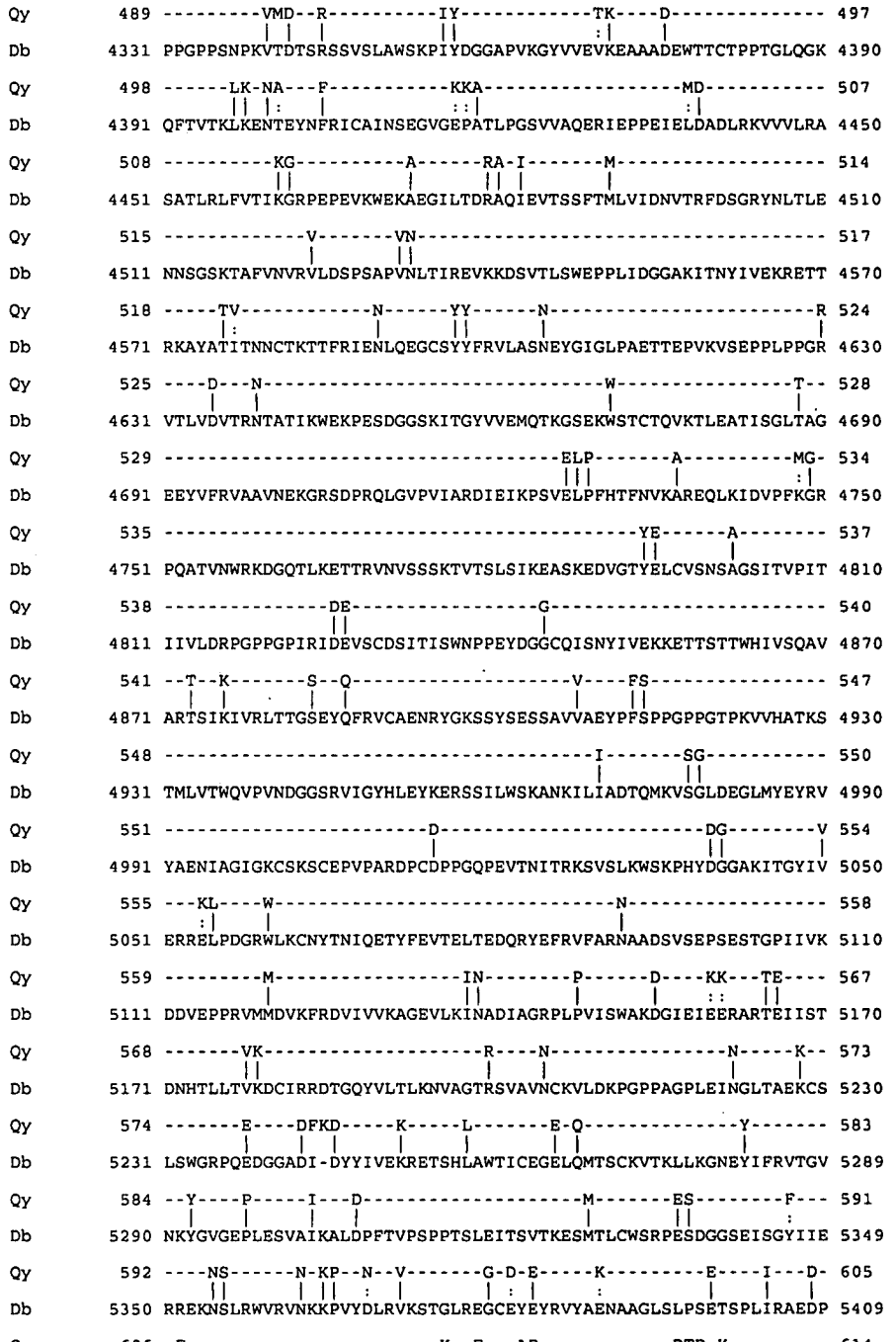
```
                     |   ||                    |||
Db       6719 SSVQNTSEPQGDKVNVSNSNSSGISNKKKYASNSLQAPGSPSRSRSATKELILPPDDSLM 6778

Qy        647 ------------------------------------KN--------I-----K------ 650
                                                   ||        |     |
Db       6779 CKPEFTKPLHDLTIHDGEQLILTCYVKGDPEPQISWSKNGKSLSSSDILDLRYKNGIATL 6838

Qy        651 -------------------S----------T---L----N---V--------I------- 656
                              |           |   |    |   |        |
Db       6839 TINEVFPEDEGVITCTATNSVGAVETKCKLTIQPLDKNINKRKVNAGDNAPKIVSHLESR 6898

Qy        657 --------N------G---------------K-S-----T-----Y------------G- 663
                   |      |               | |     |     |              |
Db       6899 FVRDGDAVNLACRIIGAQHFDVVWLHNNKEIKPSKDFQYTNEANIYRLQIAEIFPEDGGT 6958

Qy        664 Y 664
               |
Db       6959 Y 6959
```

```
RESULT 9
ABG51536
ID   ABG51536 standard; Peptide; 5701 AA.
XX
AC   ABG51536;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver peptide, SEQ ID No 30184.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00664.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human adult liver -
XX
PS   Claim 27; SEQ ID No 30184; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult
CC   liver. (I) may be used for predicting, measuring and displaying gene
CC   expression in samples derived from human adult liver. The genes
CC   identified may be involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC   is associated with coronary heart disease. ABG47348-ABG59930 represent
CC   human liver single exon encoded peptides of the invention.
CC   Note: The sequence information for this patent does not appear in the
CC   printed specification but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5701 AA;

  Query Match             69.8%;  Score 3070.2;  DB 22;  Length 5701;
  Best Local Similarity   10.3%;  Pred. No. 7e-44;
  Matches  578;  Conservative   68;  Mismatches   14;  Indels 4958;  Gaps  495;

Qy          1 Y------------------P---VV---LA-D---------T-------S---------- 9
                               |   :|   :| |         |       |
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ------------------------S-S-----E----D-----------A---------- 14
                                      | :     |    |           |
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D-----------K-E--------------- 21
                    ||     |   |          |           | |
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV--------------AEN-------------K--E-------------------- 28
                ||              |||             |  |
Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I-----------H----------------------------- 34
                   ||    : :   |           |
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA-------------M------------------------------------------ 37
                 ||             |
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 -----------ET----------------S--------------------Q--------- 41
                         ||                |                    |
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF-------------------------------------------K---- 44
                      ||                                           |
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
                  |     |                      |          |||
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE-------------------K----------E--VV--------- 57
                    |       ||                   |          |  |:
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N------------P-------V---I----D------ 64
                ||                   |            |       :   |    |
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S--------------N---------E-----------EA-KIK- 75
                    |   | | |              |         |           || |||
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ------------------------------------E-----E--------N-SNK- 81
                                                  |     |        |  ||
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT-------------------D---------------------S-F------ 90
              |  :  | ||                   |                     | |
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V-----------N---------------K-----------N---------T 95
                     |           |               |           |         :
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K----------------------E--DK-----V- 104
              |     |       ||        |                      |  ||     :
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V--------Y---I-AE--------------------FK-------- 111
```









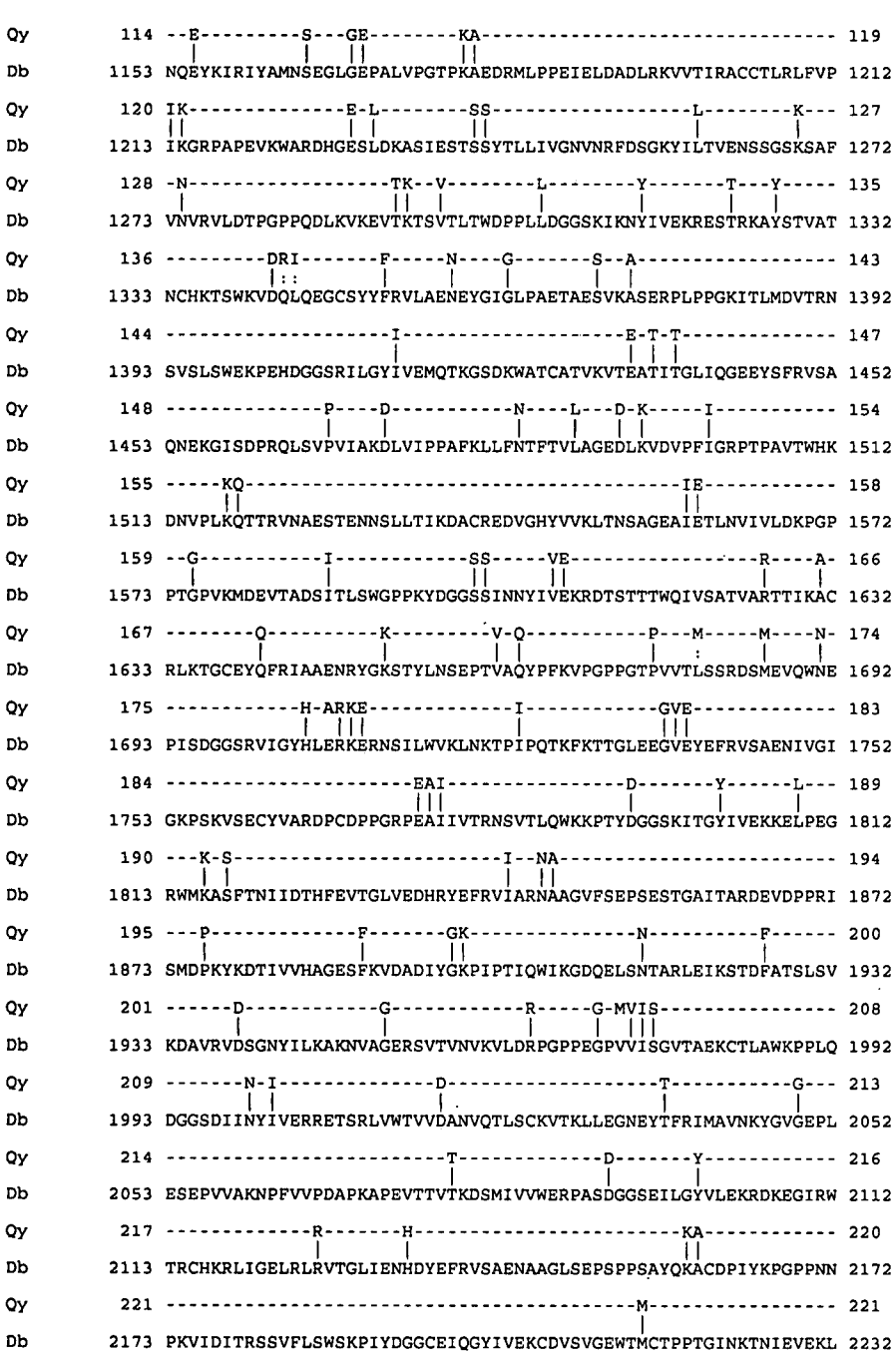
```
Db       5410 VFLPSPPSKPKIVDSGKTTITIAWVKPLFDGGAPITGYTVEYKKSDDTDWKTSIQSLRGT 5469

Qy        615 E-----L-----Y------------------------KE-------DI--------IVPA 625
Db       5470 EYTISGLTTGAEYVFRVKSVNKVGASDPSDSSDPQIAKEREEEPLFDIDSEMRKTLIVKA 5529

Qy        626 G-S-T-S------------WG-P---------------R----I------D--------- 635
Db       5530 GASFTMTVPFRGRPVPNVLWSKPDTDLRTRAYVDTTDSRTSLTIENANRNDSGKYTLTIQ 5589

Qy        636 --L----L-L--K-------P------DV---SA------P----G---KN--I------ 649
Db       5590 NVLSAASLTLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGGAPVKNYHIEKREAS 5649

Qy        650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664
Db       5650 KKAWVSVTNNCNRLSYKVTNLQEG-AIYYFRVSGENEF 5686
```

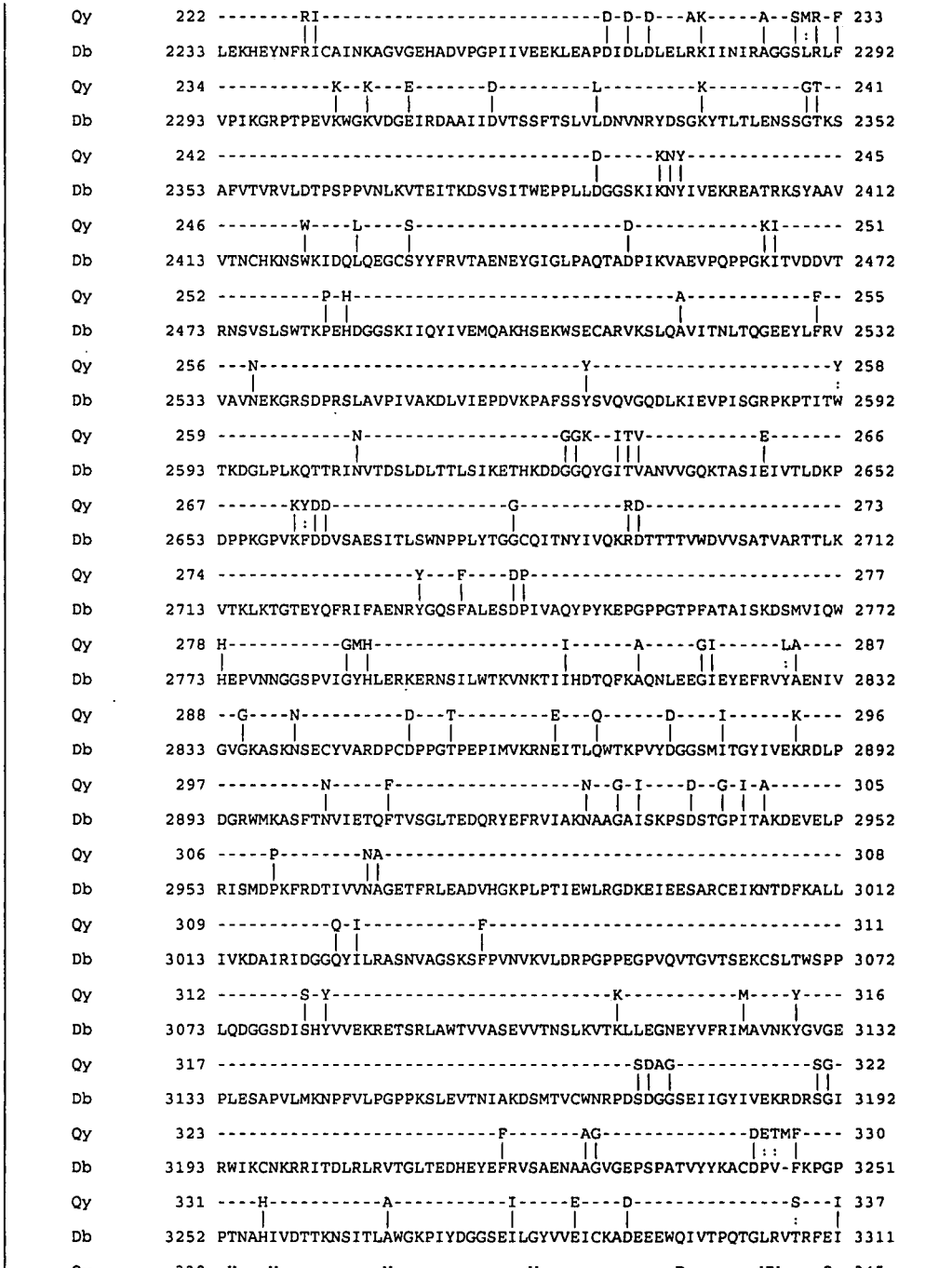
RESULT 10
ABB36684
ID ABB36684 standard; Peptide; 5701 AA.
XX
AC ABB36684;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #4190 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 29319; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5701 AA;

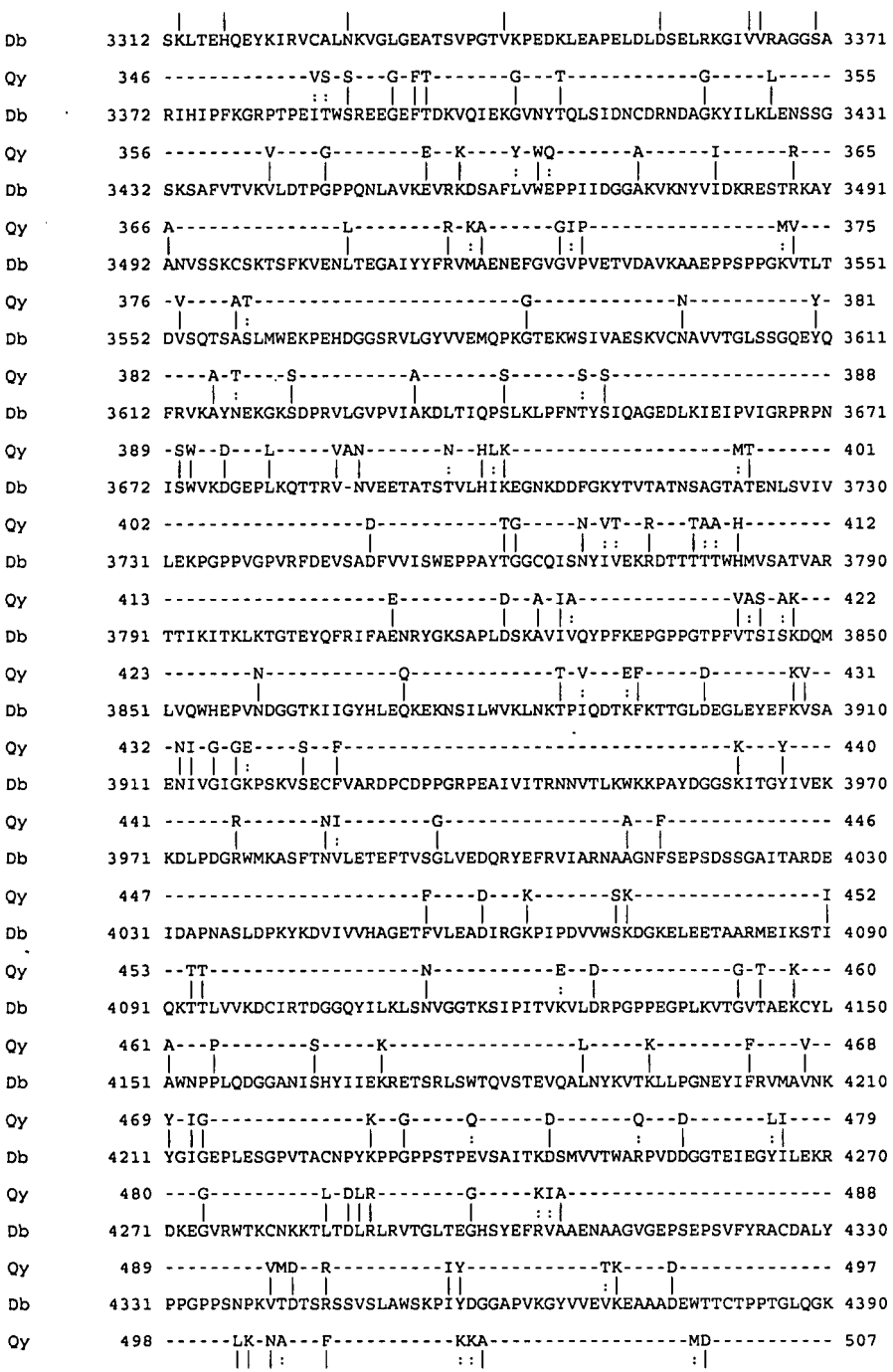
Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
Best Local Similarity 10.3%; Pred. No. 7e-44;

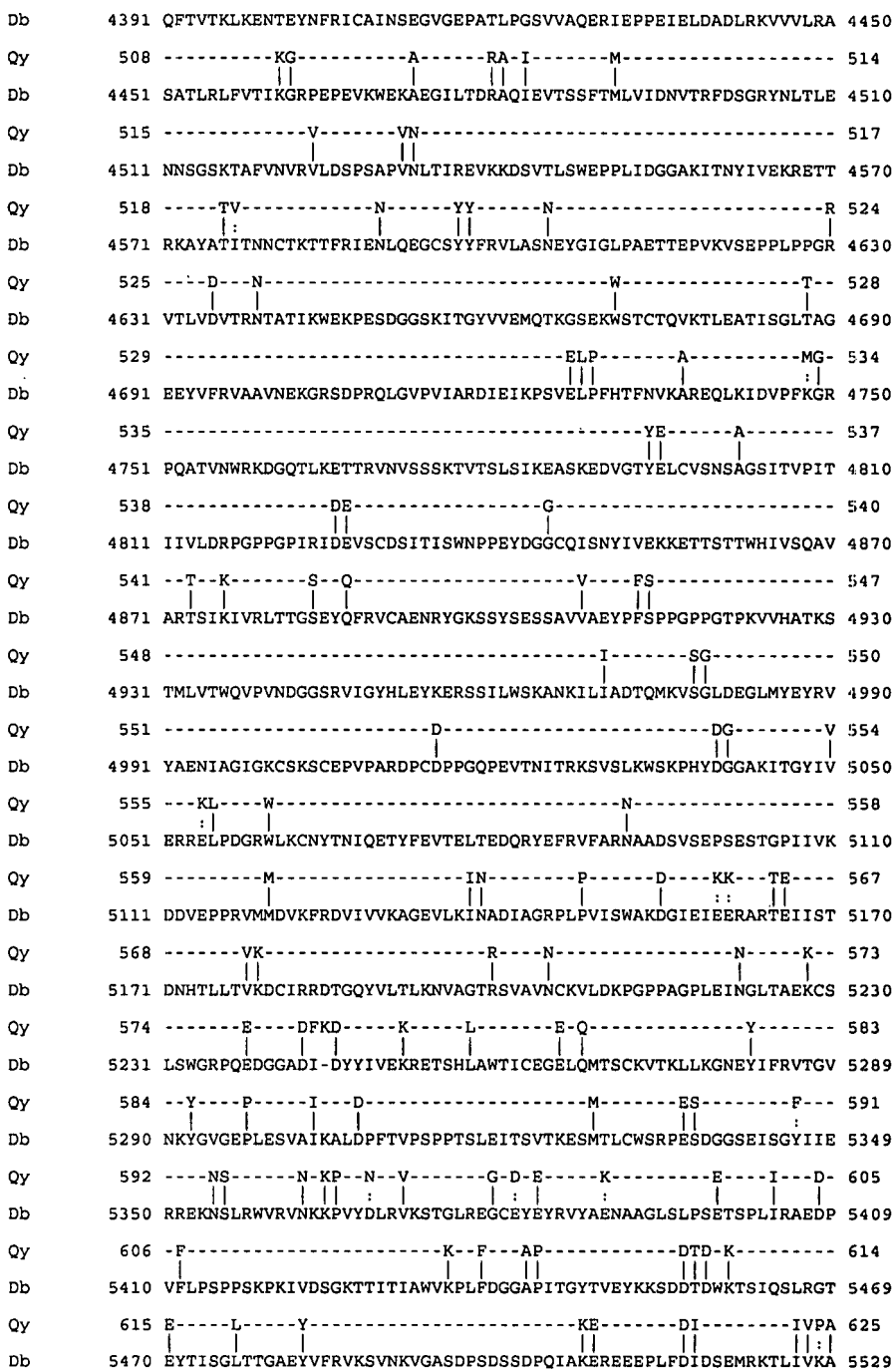
Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

```
Qy          1 Y------------------P---VV---LA-D---------T-------S---------- 9
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ------------------------S-S-----E----D-----------A---------- 14
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D-----------K-E--------------- 21
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV--------------AEN-------------K--E-------------------- 28
Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I-----------H---------------------------- 34
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA-------------M------------------------------------------ 37
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 -----------ET----------------S--------------------Q--------- 41
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF------------------------------------------K---- 44
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE-------------------K----------E--VV--------- 57
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N------------P-------V---I----D------ 64
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S--------------N---------E-----------EA-KIK- 75
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ---------------------------------------E-----E---------N-SNK- 81
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT-------------------D---------------------S-F------ 90
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V-----------N----------------K-----------N---------T 95
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K----------------------E--DK-----V- 104
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V---------Y---I-AE---------------------FK-------- 111
Db       1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFYKACDTVFKPGPPGNPR 1092

Qy        112 ------------------------------------D-----------K----------- 113
Db       1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTPPAGLKATSYTITGLTE 1152
```

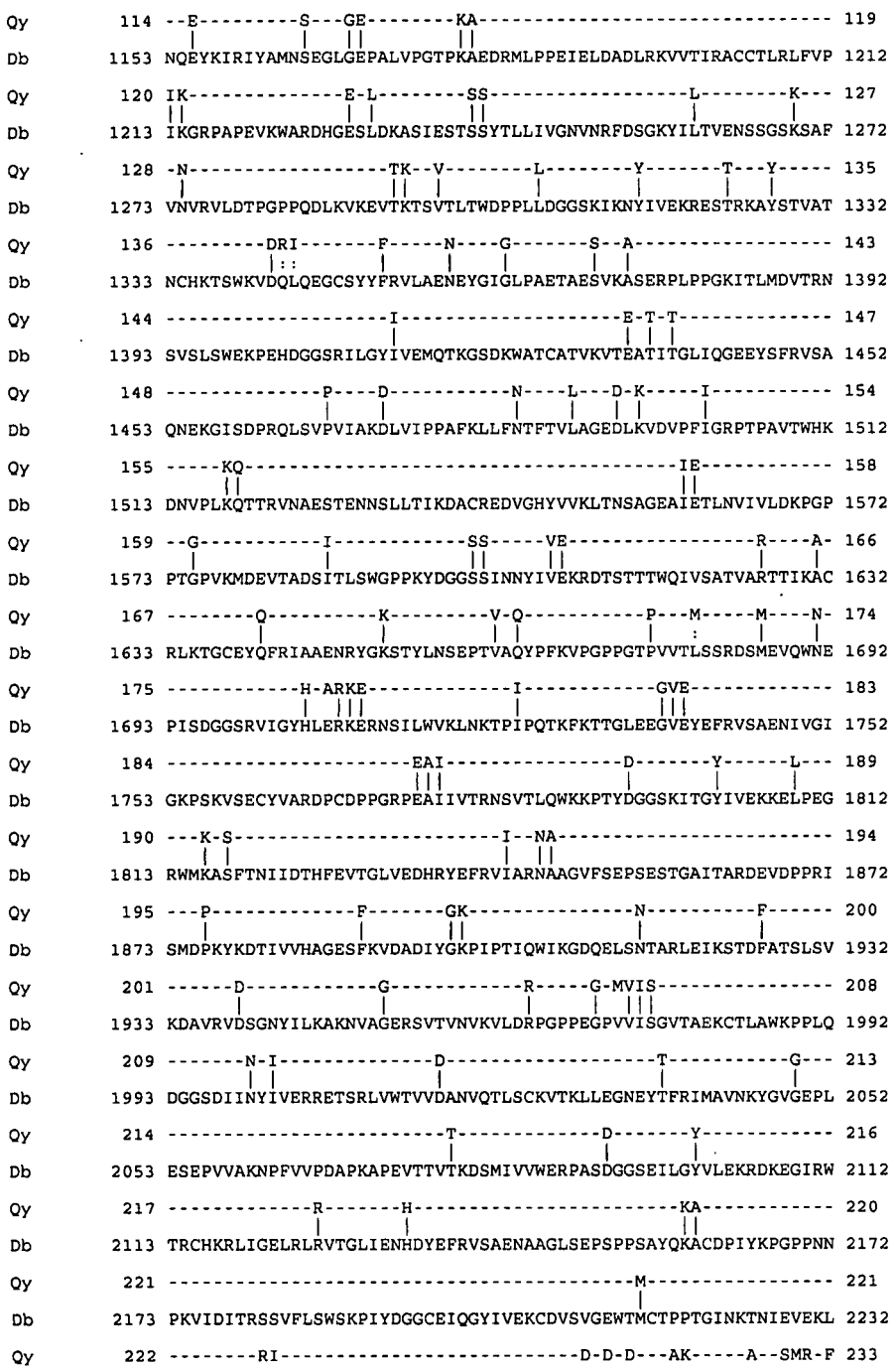


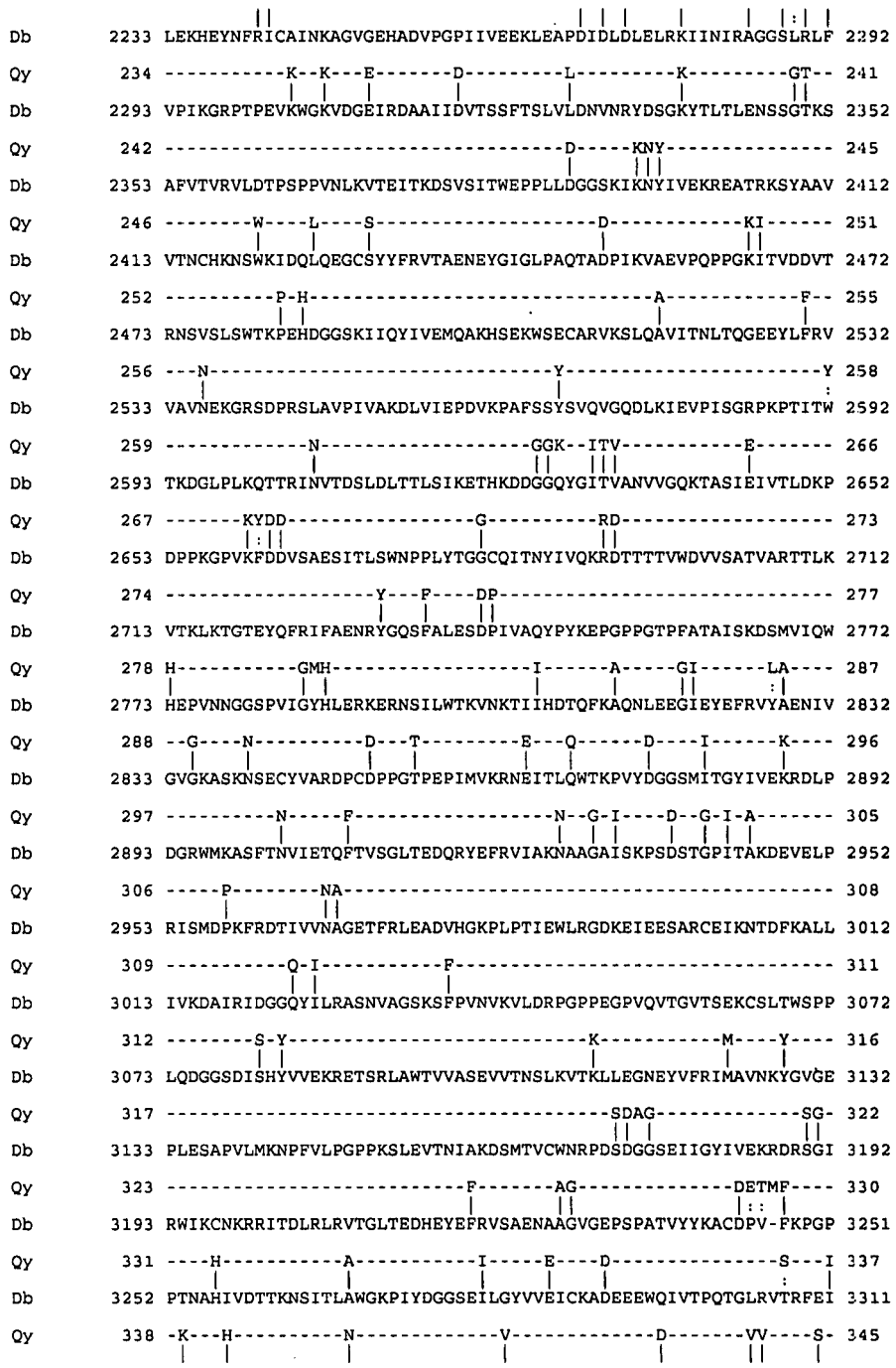


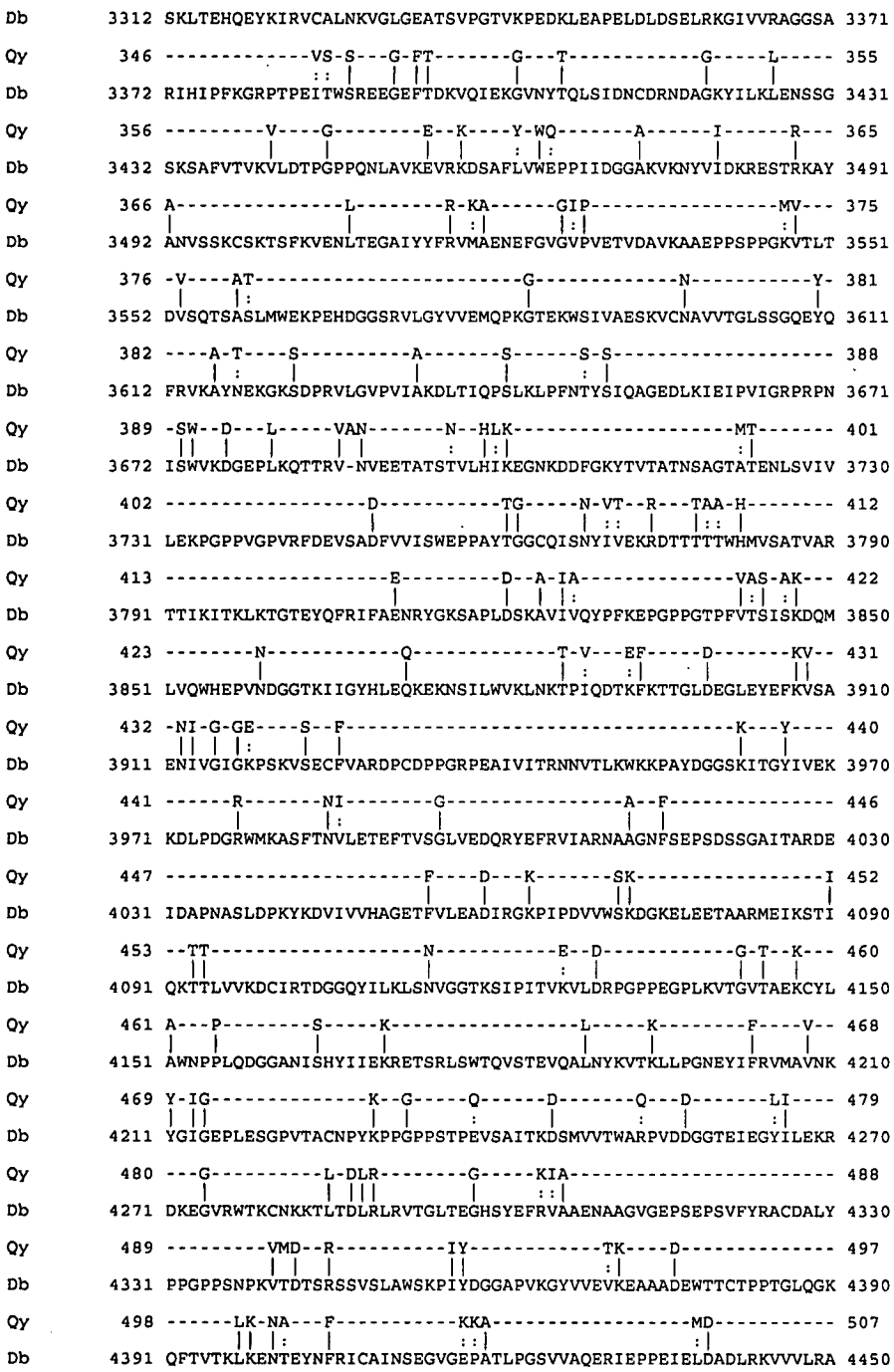


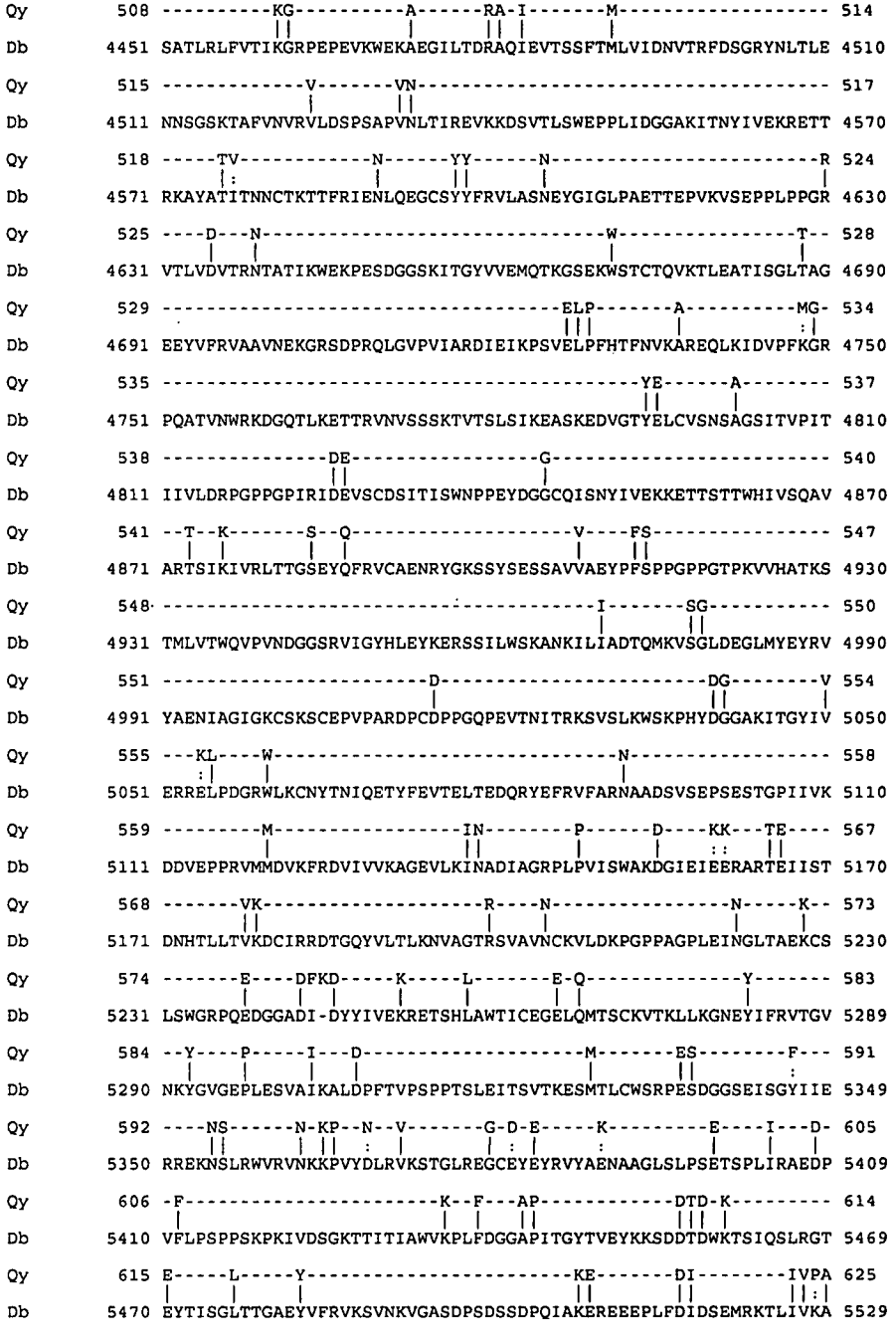


```
Qy        626 G-S-T-S------------WG-P---------------R----I------D--------- 635
              | | | :            |  |               |    |      |
Db       5530 GASFTMTVPFRGRPVPNVLWSKPDTDLRTRAYVDTTDSRTSLTIENANRNDSGKYTLTIQ 5589

Qy        636 --L----L-L--K-------P------DV---SA------P----G---KN--I------ 649
                |    | |  |       |       ||   ||      |    |   ||  |
Db       5590 NVLSAASLTLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGGAPVKNYHIEKREAS 5649

Qy        650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664
               |   | |    |     | |   | : |     |   :
Db       5650 KKAWVSVTNNCNRLSYKVTNLQEG-AIYYFRVSGENEF 5686


RESULT 11
ABB22021
ID   ABB22021 standard; Protein; 5701 AA.
XX
AC   ABB22021;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Protein #4020 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 15; SEQ ID No 23791; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5701 AA;

  Query Match             69.8%;  Score 3070.2;  DB 22;  Length 5701;
  Best Local Similarity   10.3%;  Pred. No. 7e-44;
  Matches  578;  Conservative   68;  Mismatches   14;  Indels 4958;  Gaps  495;

Qy          1 Y-----------------P---VV---LA-D---------T-------S---------- 9
              |                 |   :|   :| |         |       |
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ------------------------S-S-----E----D-----------A---------- 14
                                       | :     |    |           |
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D-----------K-E---------------- 21
                     ||     |   |          |           | |
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV--------------AEN-------------K--E--------------------- 28
                ||              |||             |  |
Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I-----------H----------------------------- 34
                   ||    : :   |           |
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA------------M------------------------------------------ 37
                ||            |
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 -----------ET----------------S--------------------Q--------- 41
                         ||                |                    |
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF-------------------------------------------K---- 44
                       ||                                           |
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
                 |     |                      |          |||
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE-------------------K----------E--VV--------- 57
                    |       ||                   |          |  |:
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N-----------P-------V---I----D------ 64
                ||                   |           |       :   |    |
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGPKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S--------------N---------E-----------EA-KIK- 75
                    |   | | |              |         |           || |||
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ------------------------------------------E-----E--------N-SNK- 81
                                                        |     |        | ||
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT-------------------D---------------------S-F------ 90
              |  :  | ||                  |                     | |
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V-----------N---------------K-----------N---------T 95
                      |           |               |           |         :
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K-----------------------E--DK-----V- 104
              |     |       ||        |                       |  ||     :
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V--------Y---I-AE---------------------FK-------- 111
                          |        |   | ||                     ||
Db       1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFYKACDTVFKPGPPGNPR 1092

Qy        112 ----------------------------------D-----------K----------- 113
                                               |           |
Db       1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTPPAGLKATSYTITGLTE 1152
```









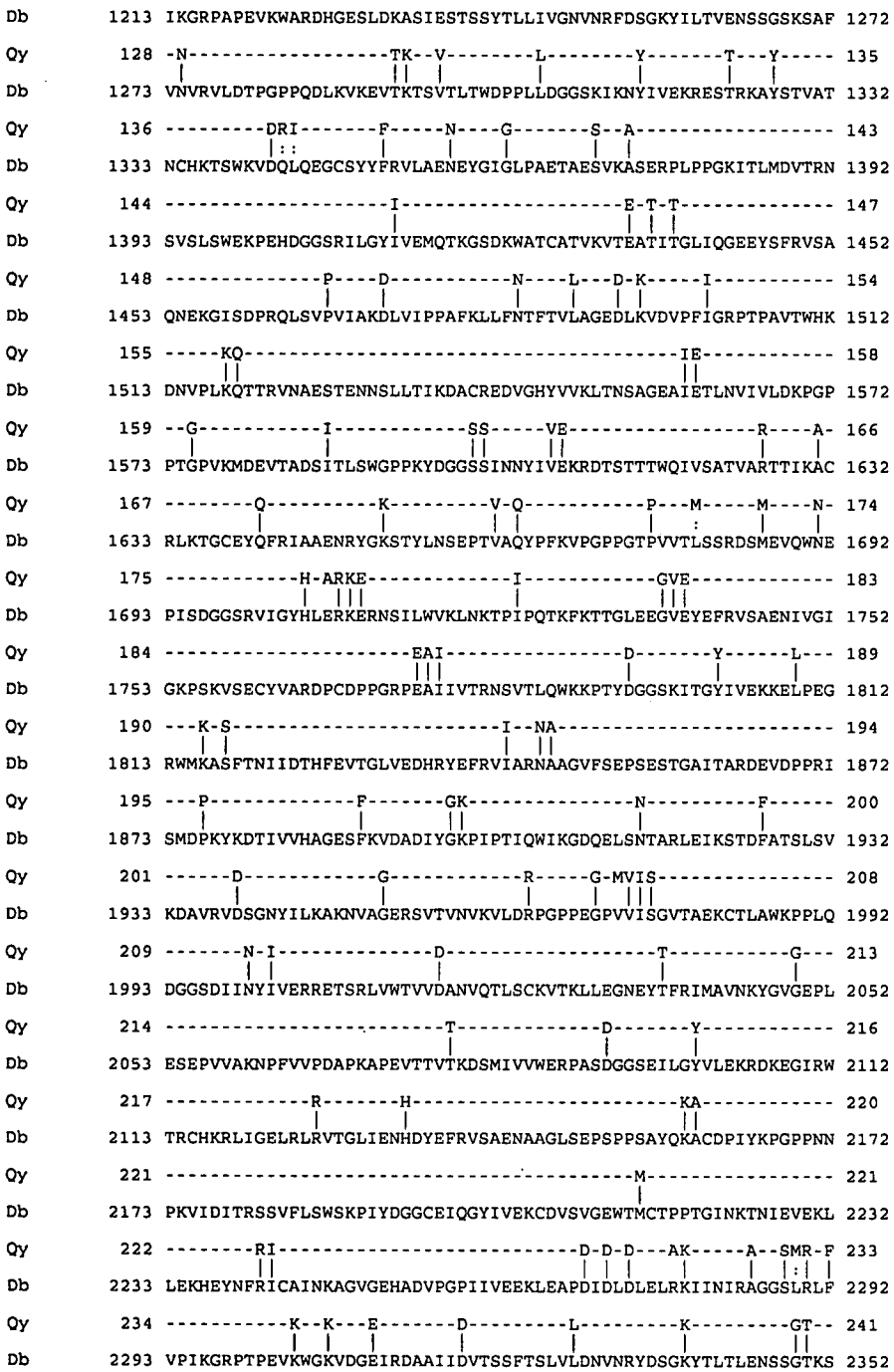
```
Qy        626 G-S-T-S------------WG-P---------------R----I------D--------- 635
              | | | :            | |               |    |      |
Db       5530 GASFTMTVPFRGRPVPNVLWSKPDTDLRTRAYVDTTDSRTSLTIENANRNDSGKYTLTIQ 5589

Qy        636 --L----L-L--K-------P------DV---SA------P----G---KN--I------ 649
                |    | |  |       |      ||   ||      |    |   ||  |
Db       5590 NVLSAASLTLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGGAPVKNYHIEKREAS 5649

Qy        650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664
               |   | |   |     | |   | : |     |   :
Db       5650 KKAWVSVTNNCNRLSYKVTNLQEG-AIYYFRVSGENEF 5686
```

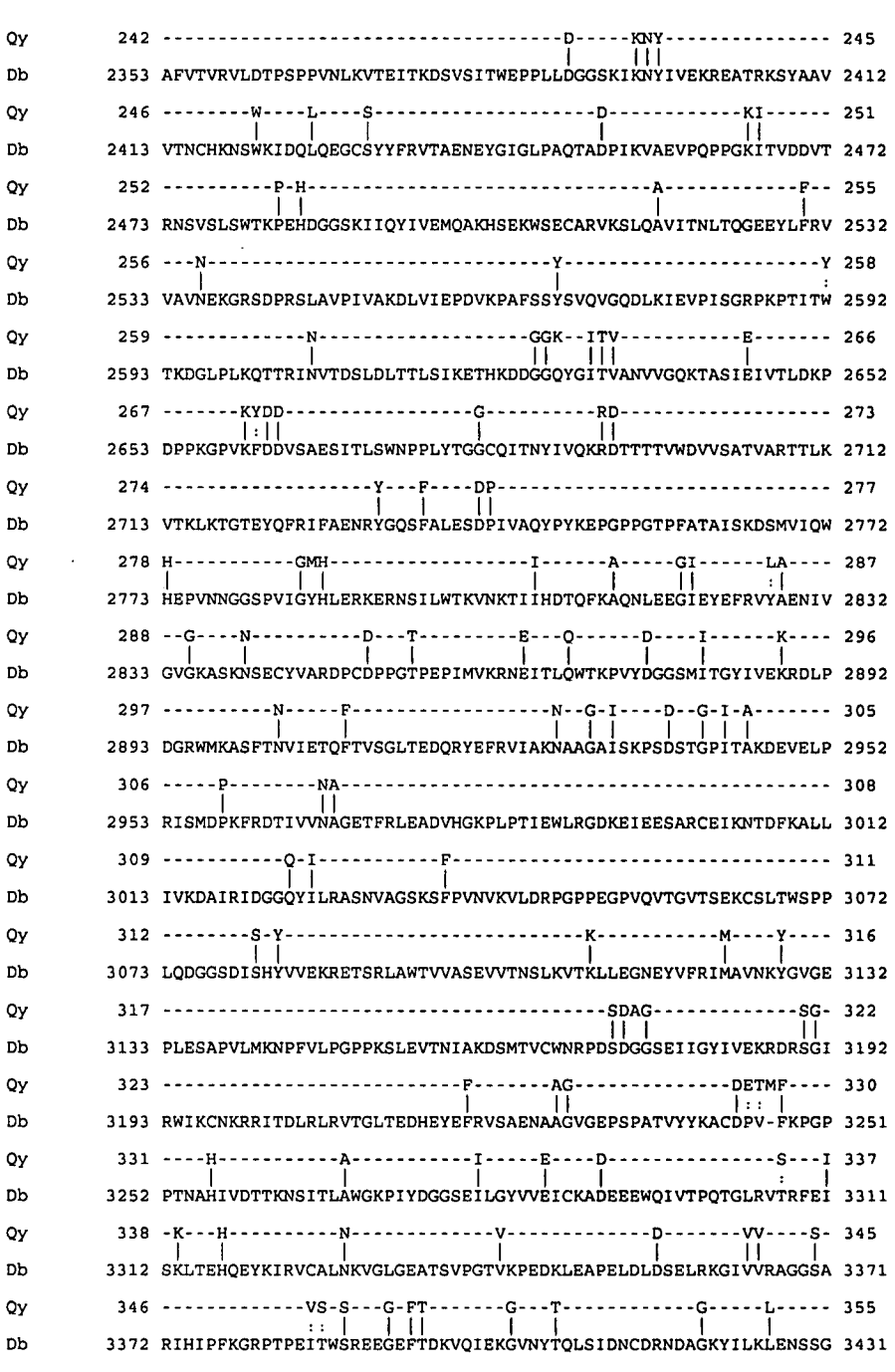
RESULT 12
AAM57448
ID AAM57448 standard; Protein; 5701 AA.
XX
AC AAM57448;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29553.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 29553; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 5701 AA;

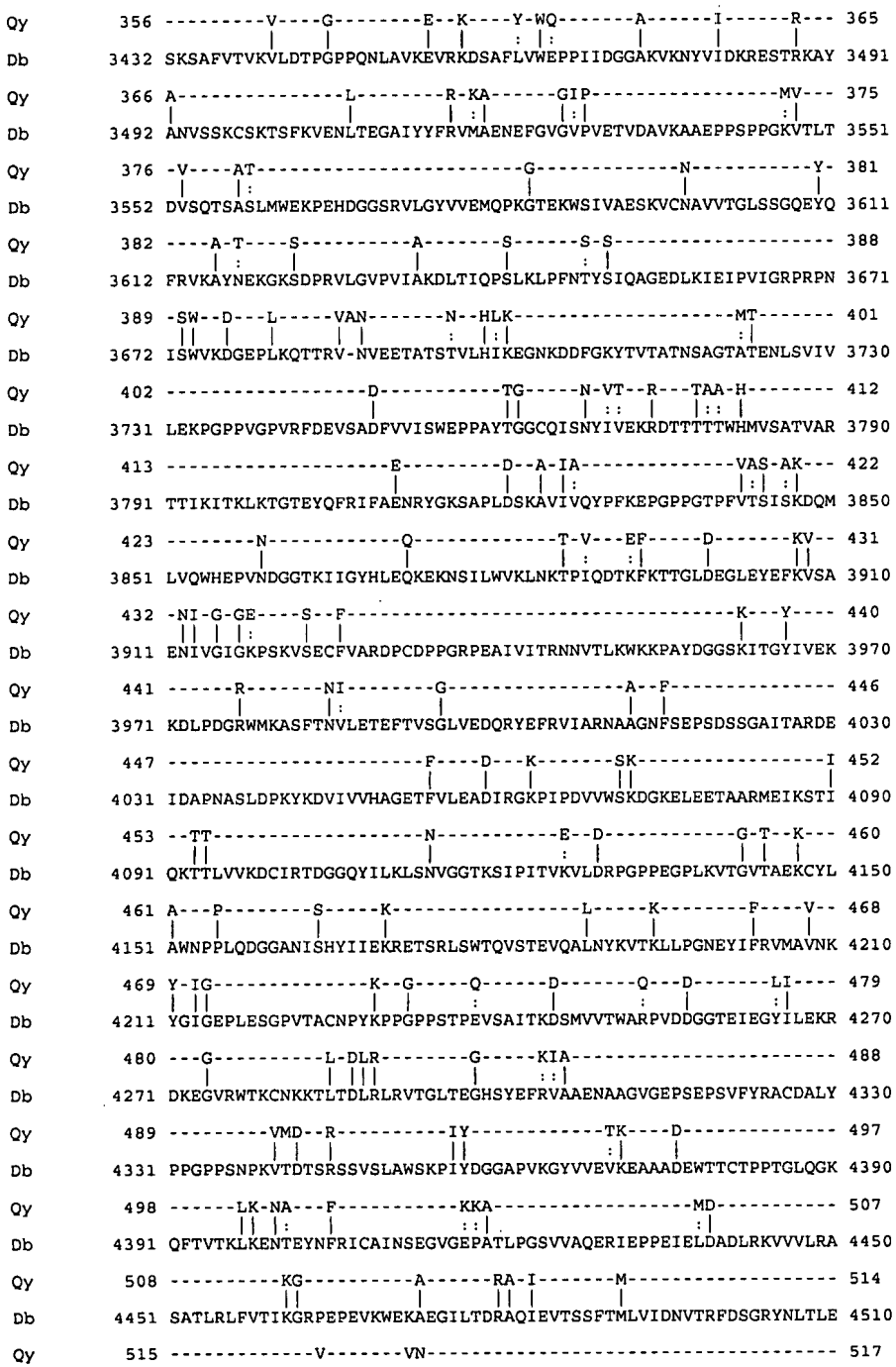
Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
Best Local Similarity 10.3%; Pred. No. 7e-44;
Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

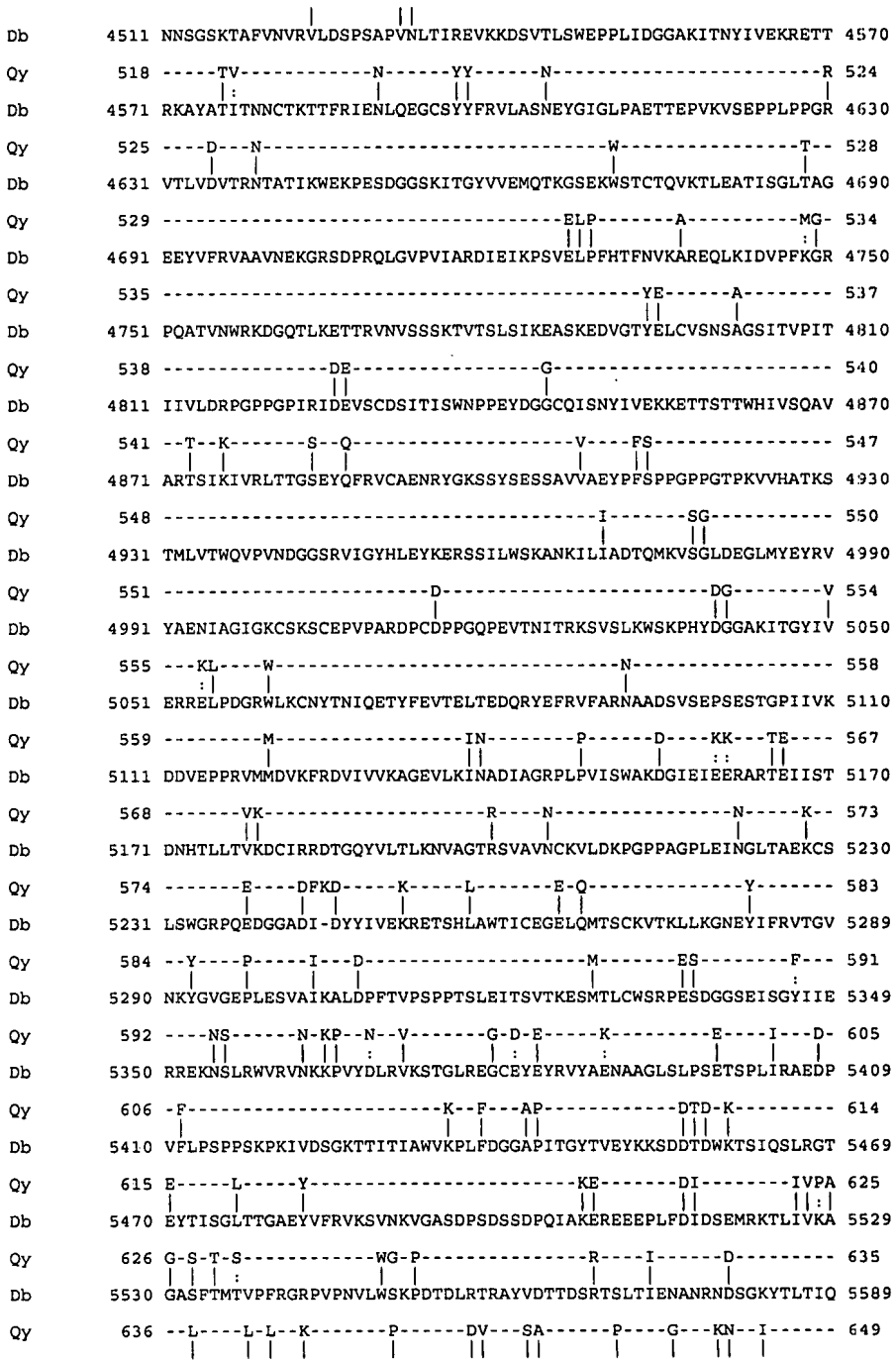
```
Qy          1 Y------------------P---VV---LA-D---------T-------S---------- 9
              |                  |   :|   :| |         |       |
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ------------------------S-S-----E----D-----------A---------- 14
```

```
                                    | :     |    |    |          |
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D-----------K-E-------------- 21
                     ||     |   |          |           | |
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV-------------AEN-------------K--E------------------- 28
                ||             |||             |  |
Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I----------H------------------------------ 34
                 ||    : :   |          |
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA------------M------------------------------------------ 37
                 ||            |
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 ----------ET---------------S--------------------Q--------- 41
                        ||               |                    |
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF-------------------------------------------K---- 44
                      ||                                           |
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
                   |     |                      |          |||
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE-------------------K----------E--VV--------- 57
                    |       ||                   |          |  |:
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N------------P--------V---I----D------ 64
                ||                   |            |        :   |    |
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S---------------N---------E-----------EA-KIK- 75
                    |   |  | |               |         |           || |||
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ----------------------------------------E-----E--------N-SNK- 81
                                                      |     |        |  ||
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT-------------------D----------------------S-F------ 90
              |  :  | ||                  |                      | |
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V-----------N----------------K-----------N---------T 95
                       |           |                |           |         :
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K----------------------E--DK-----V- 104
              |     |       ||        |                      |  ||     :
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V---------Y---I-AE---------------------FK-------- 111
                         |         |   | ||                     ||
Db       1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFYKACDTVFKPGPPGNPR 1092

Qy        112 ------------------------------------D-----------K----------- 113
                                                  |           |
Db       1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTPPAGLKATSYTITGLTE 1152

Qy        114 --E---------S---GE---------KA------------------------------- 119
                |         |   ||         ||
Db       1153 NQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADLRKVVTIRACCTLRLFVP 1212

Qy        120 IK--------------E-L---------SS-----------------L---------K--- 127
                ||              | |         ||                |         |
```

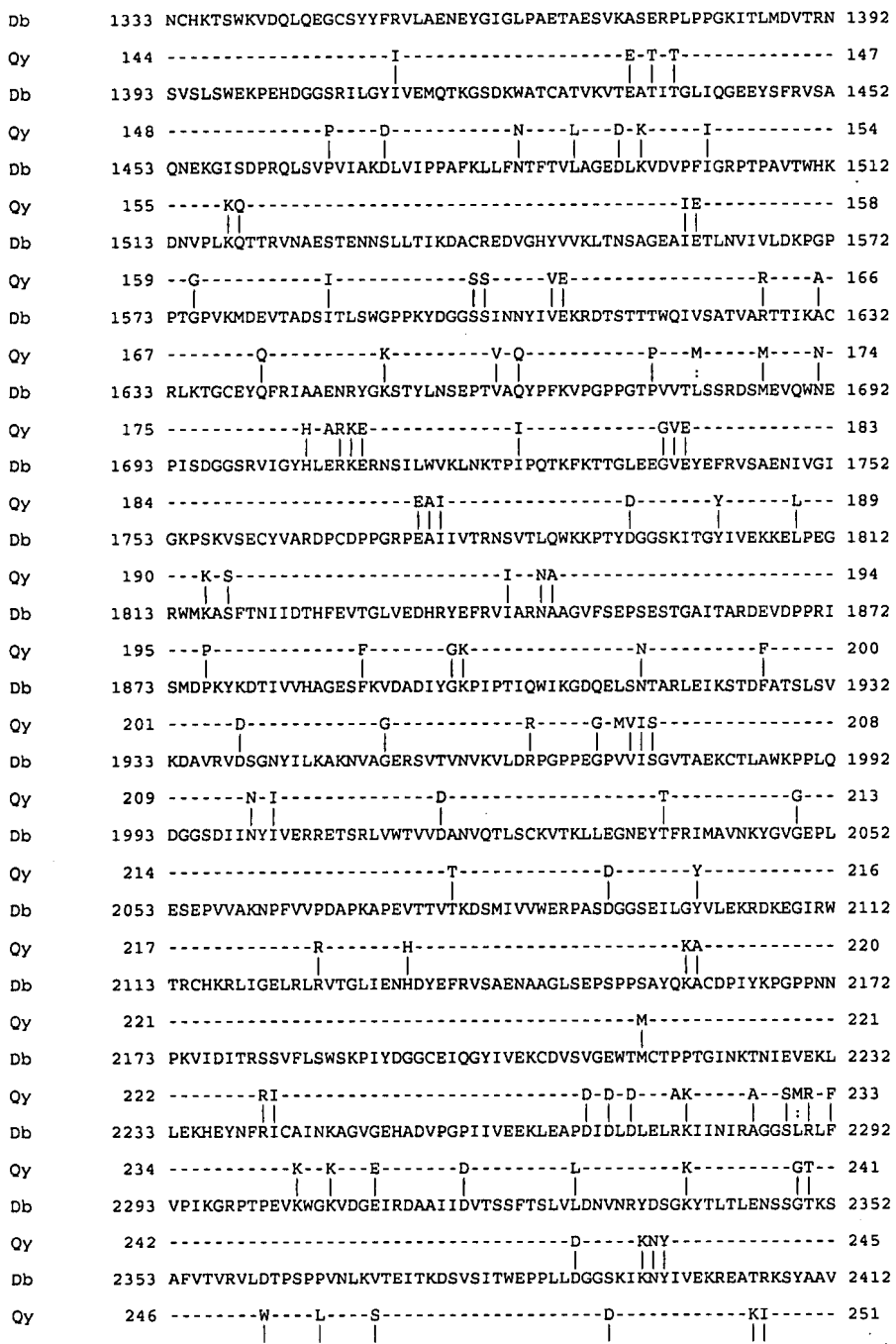


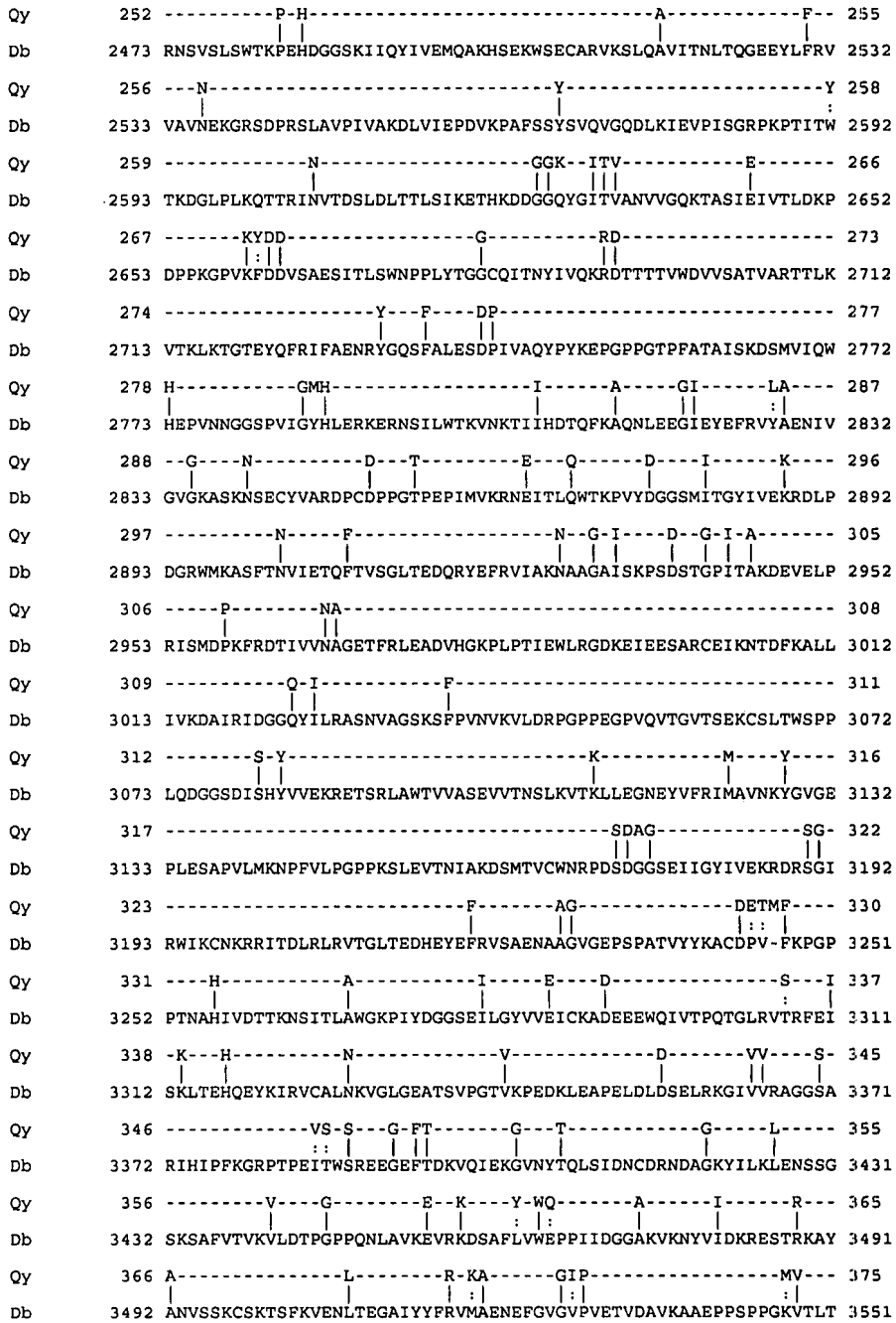


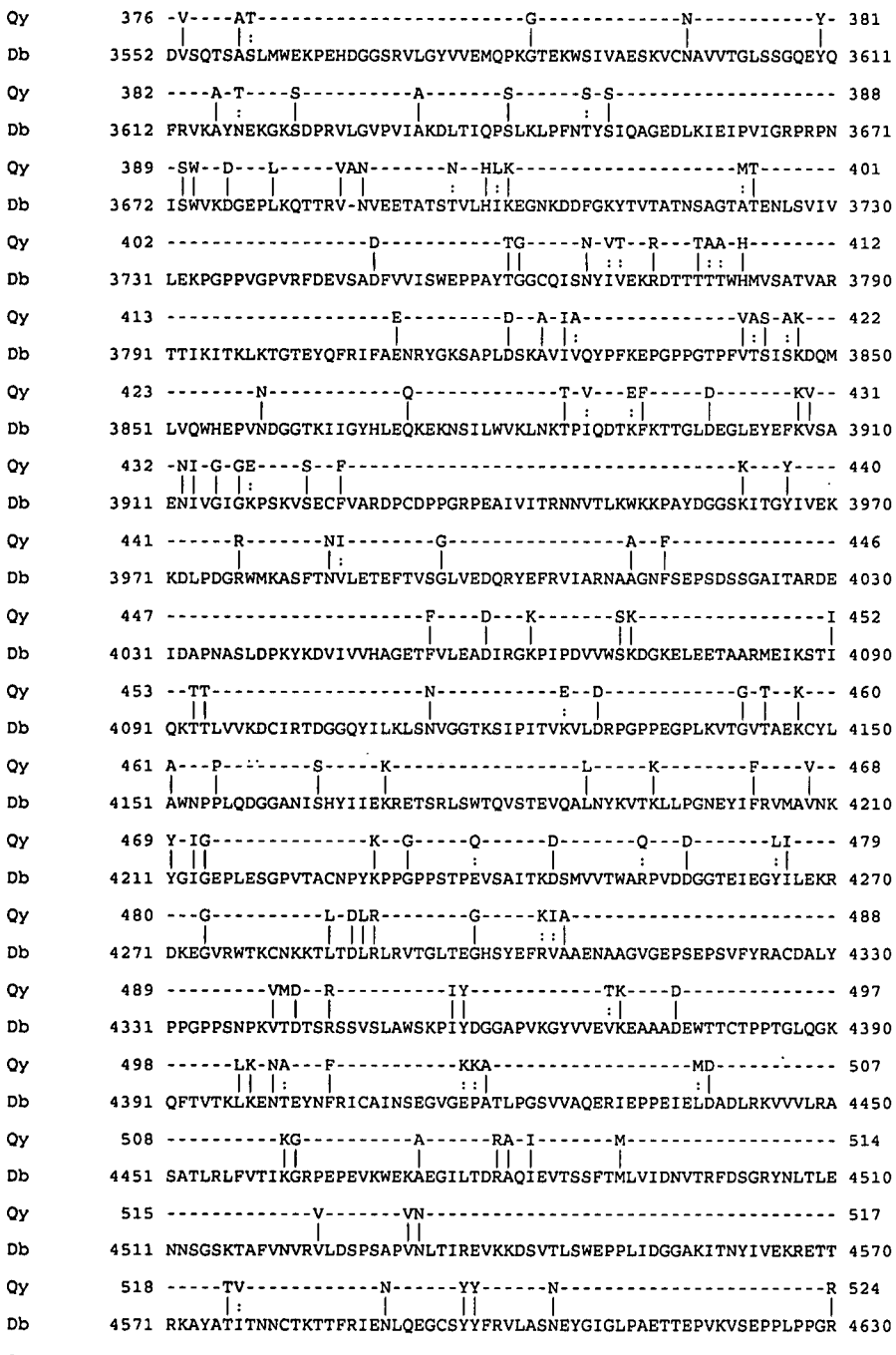


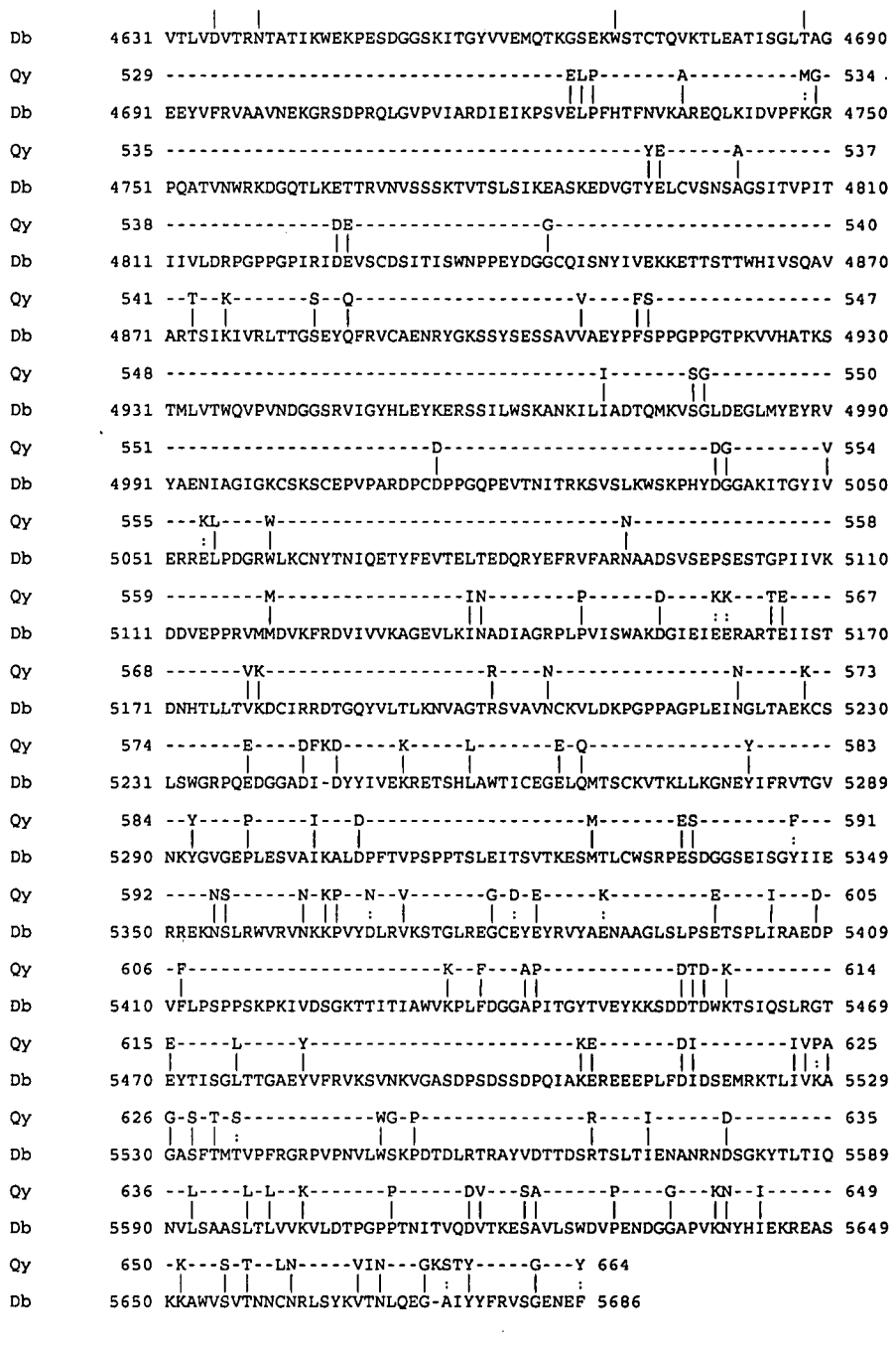


```
Db       5590 NVLSAASLTLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGGAPVKNYHIEKREAS 5649

Qy        650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664
               |   | |   |     | |   | : |     |   :
Db       5650 KKAWVSVTNNCNRLSYKVTNLQEG-AIYYFRVSGENEF 5686


RESULT 13
AAM69843
ID   AAM69843 standard; Protein; 5701 AA.
XX
AC   AAM69843;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 30149.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 30149; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   5701 AA;

  Query Match             69.8%;  Score 3070.2;  DB 22;  Length 5701;
  Best Local Similarity   10.3%;  Pred. No. 7e-44;
  Matches  578;  Conservative   68;  Mismatches   14;  Indels 4958;  Gaps  495;

Qy          1 Y------------------P---VV---LA-D---------T-------S---------- 9
               |                  |   :|   :| |         |       |
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ------------------------S-S-----E----D----------A---------- 14
                                       | :     |    |          |
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D----------K-E-------------- 21
                     ||     |   |          |          | |
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV-------------AEN-------------K--E------------------- 28
                  ||             |||             |  |
Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I----------H---------------------------- 34
                   ||    : :   |          |
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA------------M------------------------------------------ 37
                 ||            |
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 ----------ET----------------S--------------------Q--------- 41
                        ||                |                    |
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF-------------------------------------------K---- 44
                       ||                                           |
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
                   |     |                      |          |||
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE------------------K----------E--VV--------- 57
                    |       ||                  |          |  |:
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N------------P-------V---I----D------ 64
                ||                   |            |       :   |    |
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S--------------N---------E-----------EA-KIK- 75
                    |   |  | |              |         |           || |||
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ------------------------------------E-----E--------N-SNK- 81
                                                  |     |        |  ||
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT------------------D---------------------S-F------ 90
              |  :  | ||                |                     | |
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V----------N----------------K----------N---------T 95
                      |          |                |          |         :
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K----------------------E--DK-----V- 104
              |     |       ||        |                      |  ||     :
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V--------Y---I-AE---------------------FK-------- 111
                          |        |   | ||                     ||
Db       1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFYKACDTVFKPGPPGNPR 1092

Qy        112 ------------------------------------D----------K---------- 113
                                                  |          |
Db       1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTPPAGLKATSYTITGLTE 1152

Qy        114 --E---------S---GE--------KA---------------------------------- 119
                 |         |   ||        ||
Db       1153 NQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADLRKVVTIRACCTLRLFVP 1212

Qy        120 IK-------------E-L--------SS-----------------L--------K--- 127
                ||           | |        ||                 |        |
Db       1213 IKGRPAPEVKWARDHGESLDKASIESTSSYTLLIVGNVNRFDSGKYILTVENSSGSKSAF 1272

Qy        128 -N-----------------TK--V--------L--------Y-------T---Y----- 135
                |                 ||  |        |        |       |   |
Db       1273 VNVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIKNYIVEKRESTRKAYSTVAT 1332

Qy        136 ---------DRI-------F-----N----G-------S--A----------------- 143
                       |::       |     |.   |       |  |
```



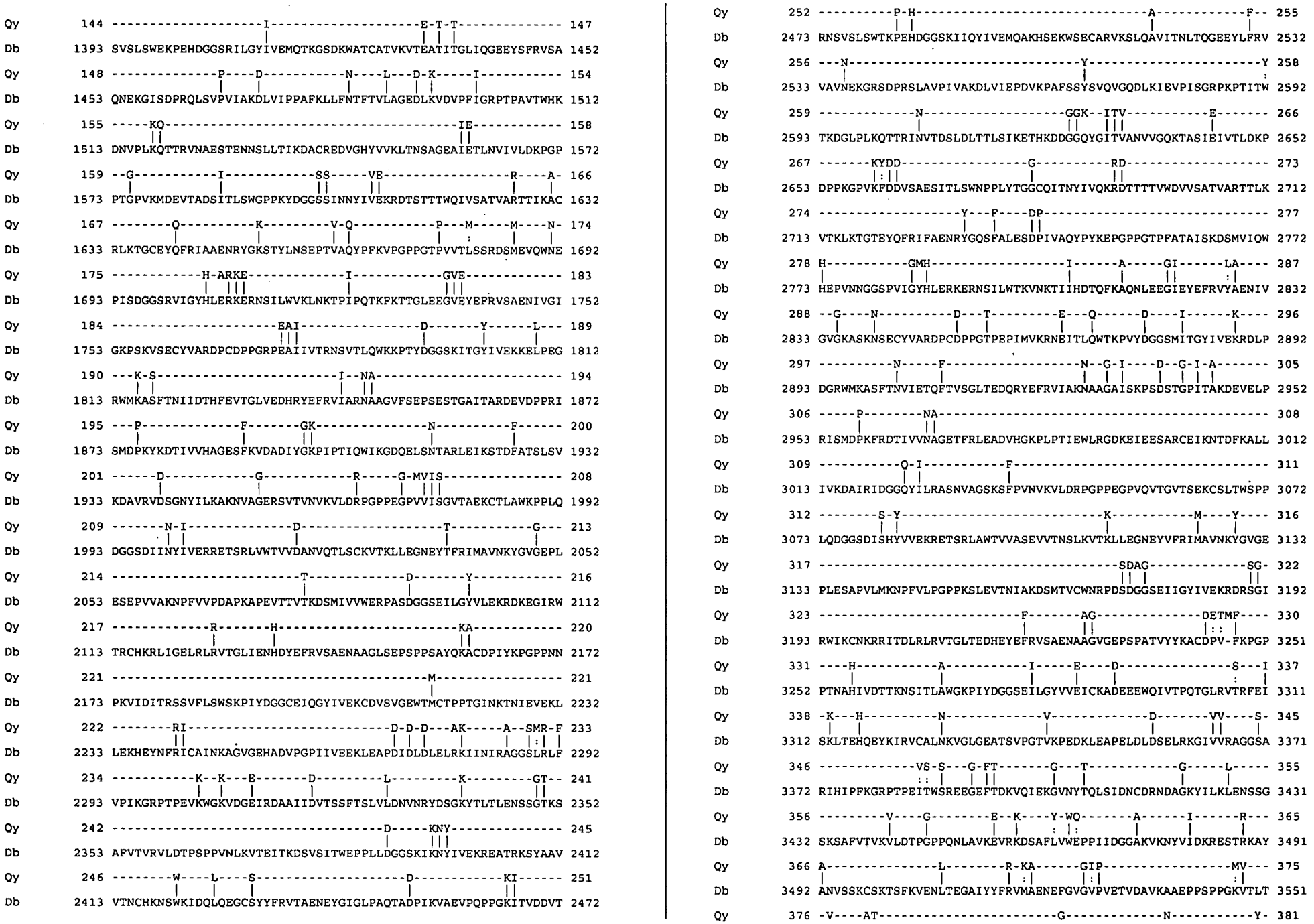


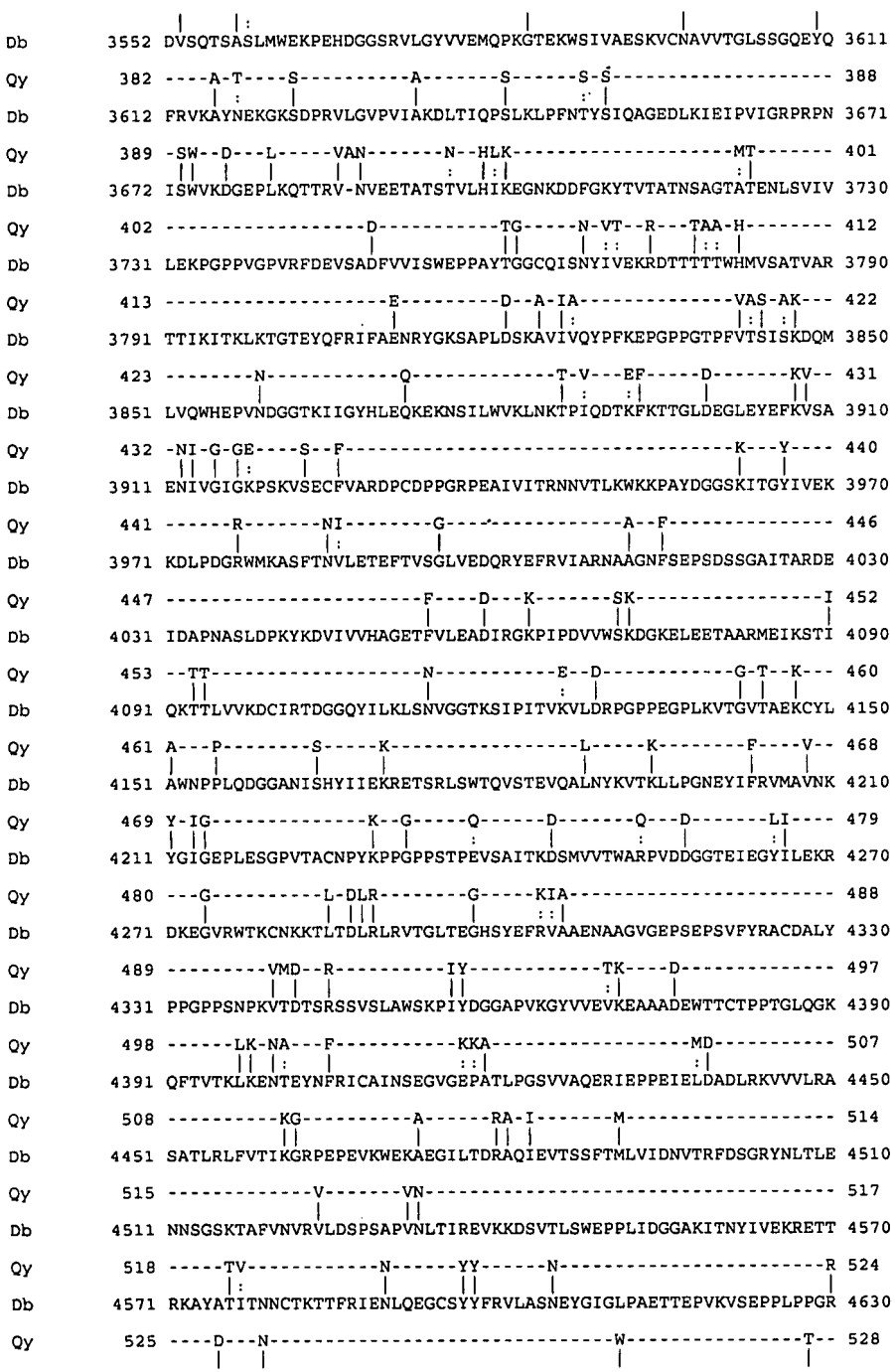


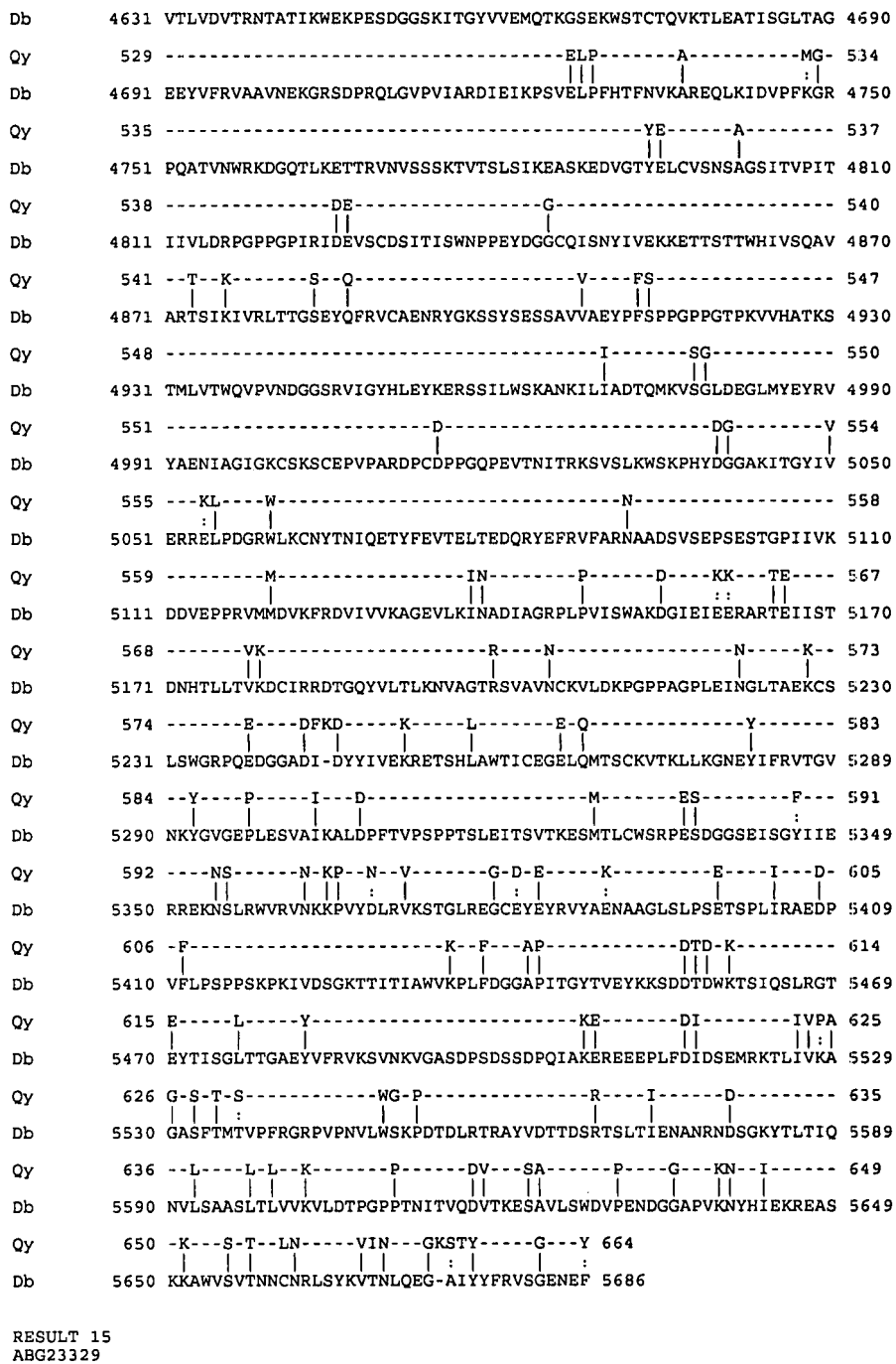


RESULT 14

```
AAM05328
ID   AAM05328 standard; Protein; 5701 AA.
XX
AC   AAM05328;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #4010 encoded by probe for measuring breast gene expression.
XX
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 27; SEQ ID No 14068; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes
CC   (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for measuring human gene expression in
CC   a human breast sample, where the probe hybridises at high stringency to a
CC   nucleic acid expressed in the human breast. The probes are useful for
CC   predicting, diagnosing, grading, staging, monitoring and prognosing
CC   diseases of the human breast, particularly those diseases with polygenic
CC   aetiology. The diseases include: breast cancer, disorders of development,
CC   inflammatory diseases of the breast, fibrocystic changes, proliferative
CC   breast disease and non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5701 AA;

  Query Match             69.8%;  Score 3070.2;  DB 22;  Length 5701;
  Best Local Similarity   10.3%;  Pred. No. 7e-44;
  Matches  578;  Conservative   68;  Mismatches   14;  Indels 4958;  Gaps  495;

Qy          1 Y-----------------P---VV---LA-D---------T-------S---------- 9
              |                 |   :|   :| |         |       |
Db         73 YNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVVRAGLSIRIFVPIKGR 132

Qy         10 ----------------------S-S-----E----D----------A---------- 14
                                    | :     |    |          |
Db        133 PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFVMTIENPAGKKSGFVNVR 192

Qy         15 --------LN-----I---S----------D----------K-E-------------- 21
                     ||     |   |          |          | |
Db        193 VLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGGSRITNYIVEKREATRKSYSTATTKCHK 252

Qy         22 ---KV-------------AEN-------------K--E------------------- 28
                ||             |||             |  |

Db        253 CTYKVTGLSEGCEYFFRVMAENEYGIGEPTETTEPVKASEAPSPPDSLNIMDITKSTVSL 312

Qy         29 -----KH----E-N---I----------H---------------------------- 34
                  ||    : :   |          |
Db        313 AWPKPKHDGGSKITGYVIEAQRKGSDQWTHITTVKGLECVVRNLTEGEEYTFQVMAVNSA 372

Qy         35 --SA------------M------------------------------------------ 37
                ||            |
Db        373 GRSAPRESRPVIVKEQTMLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGD 432

Qy         38 ----------ET---------------S--------------------Q--------- 41
                       ||               |                    |
Db        433 QILKQTQRVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492

Qy         42 ----------DF-------------------------------------------K---- 44
                      ||                                           |
Db        493 GPIKFDEVSSDFVTFSWDPPENDGGVPISNYVVEMRQTDSTTWVELATTVIRTTYKATRL 552

Qy         45 ----E-----K----------------------K----------TAV------------ 50
                |     |                      |          |||
Db        553 TTGLEYQFRVKAQNRYGVGPGITSACIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPL 612

Qy         51 ------I-------KE-------------------K----------E--VV--------- 57
                   |       ||                   |          |  |:
Db        613 SDGGSPILGYHVERKERNGILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSK 672

Qy         58 -SK--------------------N------------P-------V---I----D------ 64
                ||                   |            |       :   |    |
Db        673 PSKPSEPMLALDPIDPPGKPVPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRW 732

Qy         65 ------N---N--T-S---------------N---------E-----------EA-KIK- 75
                    |   |  | |               |         |           || |||
Db        733 LKANFSNILENEFTVSGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPKIKV 792

Qy         76 ----------------------------------------E-----E--------N-SNK- 81
                                                      |     |        | ||
Db        793 DVKFKDTVILKAGEAFRLEADVSGRPPPTMEWSKDGKELEGTAKLEIKIADFSTNLVNKD 852

Qy         82 S--Q--GDYT-------------------D---------------------S-F------ 90
              |  :  | ||                  |                     | |
Db        853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPEGPLAVTEVTSEKCVLSWFPPLDDG 912

Qy         91 --------V-----------N----------------K-----------N---------T 95
                     |           |                |           |         :
Db        913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLES 972

Qy         96 E-----N-------PK--------K----------------------E--DK-----V- 104
              |     |       ||        |                      |  ||     :
Db        973 EPVLAVNPYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWIK 1032

Qy        105 ------------V---------Y---I-AE---------------------FK-------- 111
                          |         |   | ||                     ||
Db       1033 CNKKTLTLDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPTSPFYKACDTVFKPGPPGNPR 1092

Qy        112 ------------------------------------D-----------K---------- 113
                                                  |           |
Db       1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVTPPAGLKATSYTITGLTE 1152

Qy        114 --E---------S---GE--------KA------------------------------- 119
                |         |   ||        ||
Db       1153 NQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADLRKVVTIRACCTLRLFVP 1212

Qy        120 IK--------------E-L---------SS------------------L--------K--- 127
                ||             | |         ||                 |        |
Db       1213 IKGRPAPEVKWARDHGESLDKASIESTSSYTLLIVGNVNRFDSGKYILTVENSSGSKSAF 1272

Qy        128 -N------------------TK--V--------L---------Y-------T---Y----- 135
                |                  || |        |         |       |   |
Db       1273 VNVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIKNYIVEKRESTRKAYSTVAT 1332

Qy        136 ---------DRI-------F-----N----G-------S--A------------------ 143
                       |::      |     |    |       |  |
Db       1333 NCHKTSWKVDQLQEGCSYYFRVLAENEYGIGLPAETAESVKASERPLPPGKITLMDVTRN 1392
```



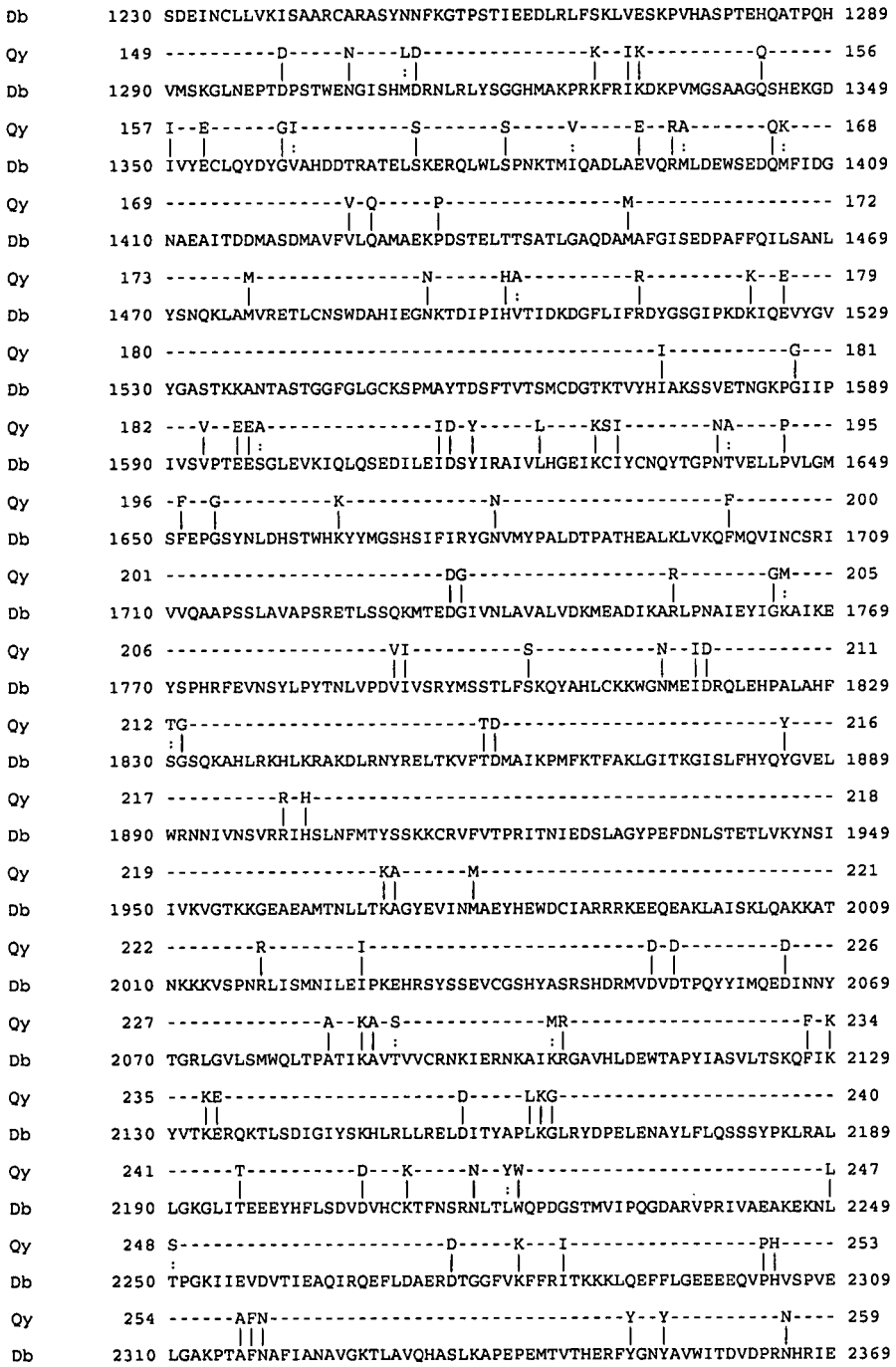


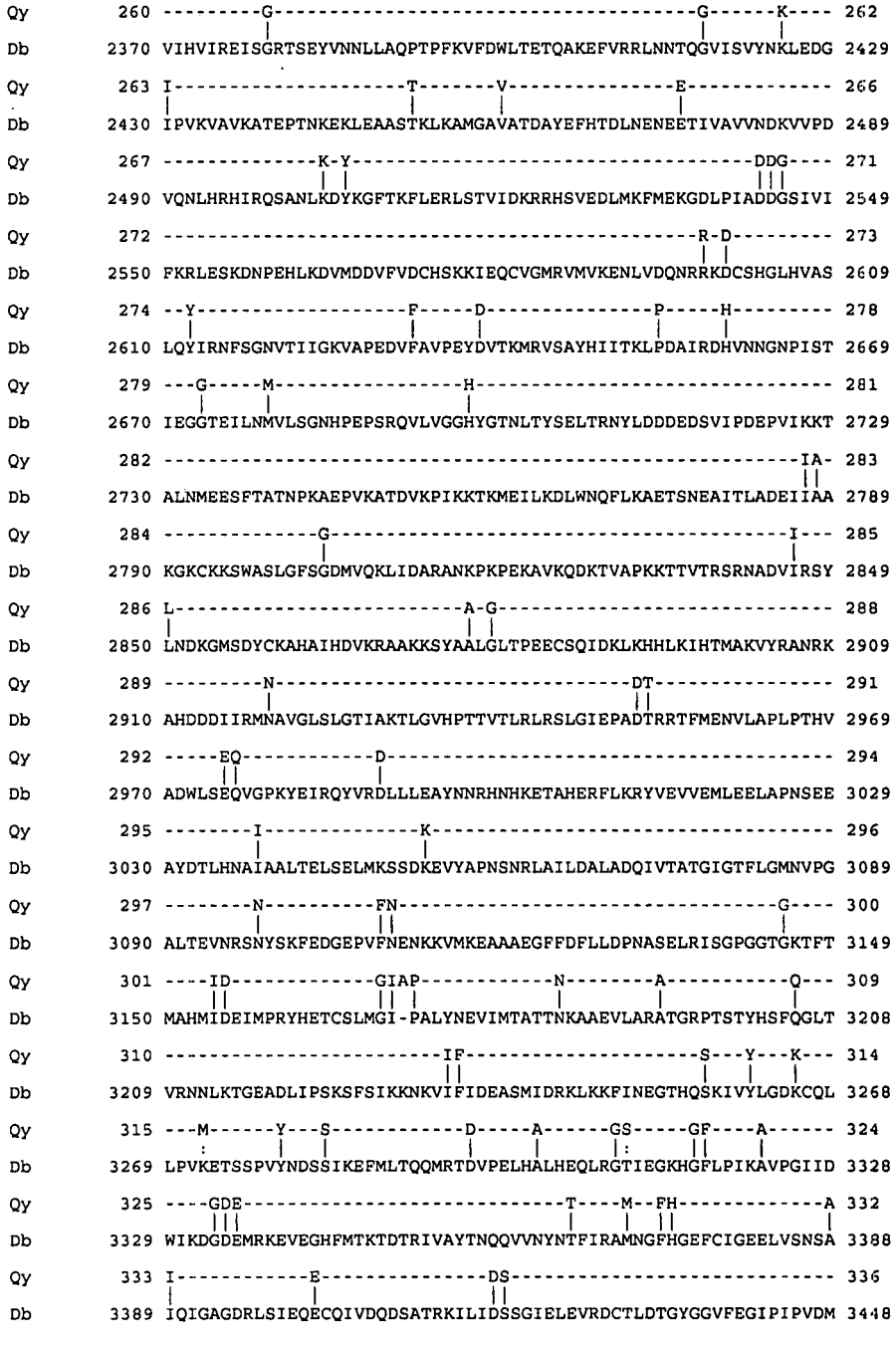


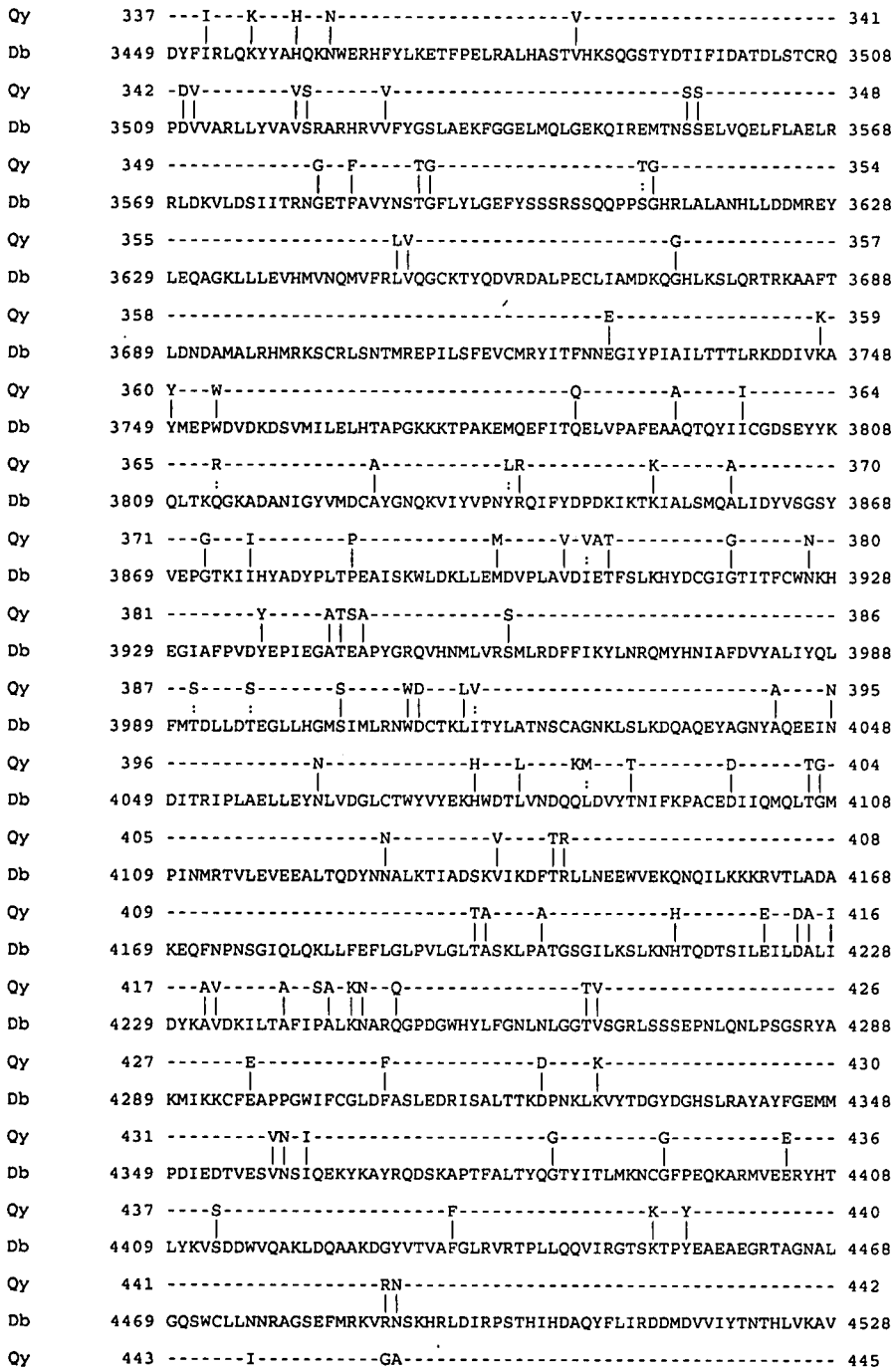
RESULT 15
ABG23329

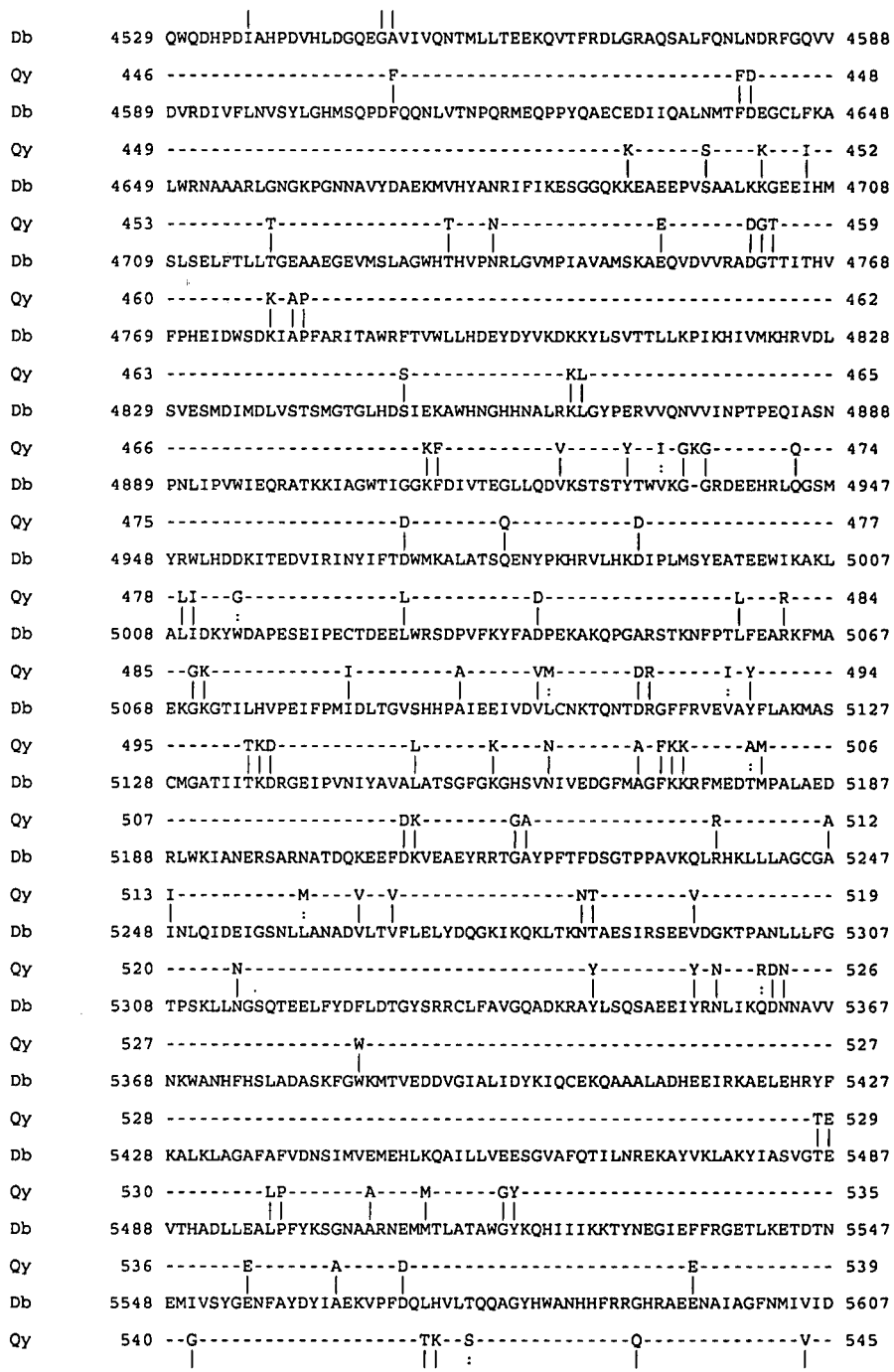
```
ID   ABG23329 standard; Protein; 6619 AA.
XX
AC   ABG23329;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23320.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87516.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 53688; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   6619 AA;

  Query Match           69.5%;  Score 3058.8;  DB 22;  Length 6619;
  Best Local Similarity   9.1%;  Pred. No. 2.8e-43;
  Matches  586;  Conservative   63;  Mismatches   12;  Indels 5802;  Gaps  489;

Qy          1 Y----P------VV--LA--------D----TS-----S--------------------  10
              :    |      :|  :|        |    ||     |
Db         30 FITSSPGMGKSAIVRQIAEDFGLKLIDHRLSTSAPEDLSGLPRFKENGRAEFAPFDELFP  89

Qy         11 ---------------------S---E----------D------------AL----N----  16
                                   |   |          |            |:    |
Db         90 LEGDAVPEGFNGWLLFLDEFNSARKEIQAAAYKLVLDRMTGQKKLHPYVAIVCAGNKATD 149

Qy         17 --I---------S-----------D--KEKVA------E----------NK----EK-H-  30
```

```
                |         |           |  :| ||      |          ||    |: |
Db        150 RAITNNLSTAMQSRLIHLEMETDFDVFMEDVAIPNKWDERVIAFLNANPNKLNDFEPDHQ 209

Qy         31 ------------------------------------------------------------  30

Db        210 EKTFCCSRTWEFVNKIVSILPPGPINSEMTLLLAGTITSGVATSFVQFTQVYSNMVSLNE 269

Qy         31 -----------E--N----------I-------HS-----A--M------------ET--S  40
                       |  |          |       ||     :  |            :|  |
Db        270 ILDNPKKARMPEDNNLLWAVVTSLINNTDEDNHSKIFDYVERMPFTFKVLYYRSRKTDMS 329

Qy         41 ----Q-----D------F--K---------E--K----K----TA---------------  49
                |     |      |  |         :  |    |    ||
Db        330 MTDAQLTREYDRCQAKAFMGKTAAFFGSLLCSLKFRWVKDGCDTAQTDGEHLEWNPDWFE 389

Qy         50 ---------VIK-E----------------KEV-------------V----S----KN--  60
                      |:: |               :||             :    |    :|
Db        390 SLLPESRVTVLMHELWHVGLLHSVRMGSRDPEVWNYACDIYINNQLIQDGYSFIGIENCW 449

Qy         61 --P---------V--------I--------------D-------------NN------  66
                |         :        |              |             ||
Db        450 KDPKYAGWAEEQIYDDLMSKQIRPPKASGAFGTGSGGDMKPSTSKASQVNIVNNVVRAMH 509

Qy         67 ------------------------------------------T--S----N-------  69
                                                        |  |    |
Db        510 QQKLSGGSLPGVLPGKMEEVITQFLKPVVPWQVLLERFFNDLQETYYSWQRPNRRYPDMY 569

Qy         70 -----------E----------------------E-A-----------------------  72
                       |                        | |
Db        570 LPSPMDDDGRLEHLAYFLDTSGSIKPKDALRFSSEVAYVKEKYQPQKMTLVQFTTEIVDE 629

Qy         73 --------------K--------------IK------------E----E-----------  77
                           |              ||            |    |
Db        630 QVIEDGDKFTDVTIKGRGGTSLVPVREWIIKHKPTAAIIFTDLEVRPMEPLPFDIPVIWV 689

Qy         78 ---N--S---------N-K--S-----------------Q--------G--------D--  85
                |  :         | |  |                 |        |        |
Db        690 VIRNPGAQVPFGKGMVNSKLSSGETTFGLGEAGYDIRIKQDVTFYRFLGLIPCIKIKDEG 749

Qy         86 ---Y-----T-------------------D---------S-F-----------------  90
                :     |                   |         | |
Db        750 KTKWKFGTFTLASAMERFNVPAGLVGIVHDKSTWARRTLSVFNTVIEPGWSGYLTLELVY 809

Qy         91 ----------------VN--K------N----TEN-----------------------  97
                           || |      |    |||
Db        810 HGRKKLHIPAGSGIAQVNGQKVTVVARNVAYVTENSEGRGVITFTNGDSIDTQVGYDSIR 869

Qy         98 ------------------PK---K--ED-----------------K--------VV---  105
                             ||   |  ||                   |        ||
Db        870 RNVAKALDGAKDIAEKRTPKLYQKTSEDGLALNQKAIKAVSALVQGKQDAVALRLVVEQI 929

Qy        106 ----------------------------Y----------------I------------  107
                                         |                |
Db        930 LVALTLAQVAKQEGKQETCKTALGALYWLYKDLADNFEMSMDKWQGISCVLALREGQELL 989

Qy        108 -----AEFK-----------------------------------------D----KE--  114
                   ||||                                        |    .:|
Db        990 STLPKAEFKLAYTGAHKIVLGSIRTSLTTSSYSRLSMINAKILADSITLSGDRLVTMELV 1049

Qy        115 ---------------------S----------------------G--EK---A-------  119
                                  |                      |  :|   |
Db       1050 YPRFIHSEFMTHRALNRNASSSRAIPTQRLIEQVRNEPVMPVHWGKNQKGMQASEELTPM 1109

Qy        120 -IK--EL------SS--------L-------K-----------------------N------  128
                ||  |:      |:       |       |                       |
Db       1110 EIKDAEFIWENAASAAAVYAEQLRRGQVHKQIVNRILEPFTHIRVVVTATQWNNFFGLRI 1169

Qy        129 ------------TK------------VL--------Y-T--------YD-----RI------  138
                         ||            :|        | |        ||     ||
Db       1170 HPDAQPEICELATKMKEQYDQSKPRTLLAGEWHLPYITNSDYLDAYDYCKHQRITRDEPS 1229

Qy        139 ----------------------FNG--SAIE-------------------T----TP--  148
                                   | |  |:||                     |    ||
```









```
Db       5608 VDGTCSLDVCHELMKEYRFMTYTTKRHTDEENRFRLIIPMNYQLELDGAEYKEFMDNVMA 5667

Qy        546 ---F----SI--------S---G----------D--D------------G---V---- 554

Db       5668 WLPFKTDESANQRAKKWESCAKGTYYYNMDAPLLDVRDFIPRTSKNEAFQQGMQKVQNLD 5727

Qy        555 KL--W----------N-MI---------------------------NP------D--- 563

Db       5728 NLERWFAQRIATGNRNNHMIKYALALVDSGIGFNEVKQRVHEFNKKLSNPLSADELDSTV 5787

Qy        564 -----K----------K--T--------E--------V------K-------RN----- 571

Db       5788 MITVAKLLFLVEEDCQLKEETMSEEFEHGEMNDQLILIVGFSAAGKSASLRNIRNQERWV 5847

Qy        572 --N----K-------------------E----------------D---F-KDKLE-QY 583

Db       5848 YLNTEAGKRLPFKNNFNGVRITDPYQILEYFDQCIEHRDDVDGIIVDSLTFMMDMLETQY 5907

Qy        584 ----------Y--------------------P-I---------D---ME--SF------- 591

Db       5908 VLTAANTQKAWGEFAQFFKILLQEKVVKFAKPVIFTAHVKDEVDERAMELKTFVPVKGSL 5967

Qy        592 --N----------------------SN--------------K------P---NVG---- 599

Db       5968 KNNGIEAYFSTVVSAERIDLKELEKYSNGMLEITEDEQELGYKHVFQTRPTKKTVGKRIR 6027

Qy        600 ---------------------D---E--------K-E--ID-----F------------ 606

Db       6028 SPMGMFSKQETYMDNDAQKLLDHLAEYYSTGQAAKAEKATDNLGGGFGAKESDIYLANLK 6087

Qy        607 -----K------F----APD-TD---------KE-LY----------------KE---- 619

Db       6088 VAYAGKAASGANFIQIIA-DLTDLDGHSAGEYREQLYITSGTEKGCKCTYEKNGKEYFLP 6146

Qy        620 ------DI--------I-------------------------VP--A--G--------S 627

Db       6147 GYTVINDILVMTSGETIPEAVFEEKVVNVYDFDEKKEVAKSVMVPVNAIGGKFAVAILKS 6206

Qy        628 -----T--------S------------------------------------------- 629

Db       6207 EEDKQTKDGSGNYVSTGETRFTNTIEKVFHPDLHLTVVEAEELTERGKELTVEEAVFWDK 6266

Qy        630 W-------------------------------------------------G--P---- 632

Db       6267 WLEKNKGVTRDKTTKGGASGKAGQPPKPGATNTGAGASAAKSLFEIMQIPVLGVDPSFRN 6326

Qy        633 ----R--------I----DL-LL--KPD----V---S-----------------A--- 644

Db       6327 WGLARGMLDLETGILSGLDLKLVETKPDGTKQVRQNSKDMQAAEDITTGVIDWFKEAKVI 6386

Qy        645 ----P-G---KN-IKS----------T-LN-----V--I-N-----G-K--S------- 660

Db       6387 FVEVPVGSQSANGMKSYGVCVGILGALTALGHEIIEVTPIENKVALSGIKTASKDVMIRA 6446

Qy        661 -T--Y-----------------------------------GY 664

Db       6447 ATGFYPEANWLKDTKEFNENFIKGITMQVILTQPEIEAALTGY 6489
```

Search completed: October 9, 2003, 15:57:54
Job time : 118 secs